Globally Convergent Homotopy Algorithms for Nonlinear Systems of Equations

By Layne T. Watson

TR 90-26

GLOBALLY CONVERGENT HOMOTOPY ALGORITHMS FOR NONLINEAR SYSTEMS OF EQUATIONS

Layne T. Watson†

Abstract. Probability-one homotopy methods are a class of algorithms for solving nonlinear systems of equations that are accurate, robust, and converge from an arbitrary starting point almost surely. These new globally convergent homotopy techniques have been successfully applied to solve Brouwer fixed point problems, polynomial systems of equations, constrained and unconstrained optimization problems, discretizations of nonlinear two-point boundary value problems based on shooting, finite differences, collocation, and finite elements, and finite difference, collocation, and Galerkin approximations to nonlinear partial differential equations. This paper introduces, in a tutorial fashion, the theory of globally convergent homotopy algorithms, describes some computer algorithms and mathematical software, and presents several nontrivial engineering applications.

1. Introduction.

Continuation in various forms has been used for a long time in mathematics and engineering, with such names as parameter continuation, incremental loading, displacement incrementation, imbedding, invariant imbedding, continuous Newton, and homotopy. The state-of-the-art of continuation methods was thoroughly surveyed in [1], and more recently in [74]. Recent mathematical developments have led to a whole new class of continuation methods known as *probability-one homotopy algorithms*, which have been successfully applied to solve Brouwer fixed point problems, polynomial systems of equations, and discretizations of nonlinear two point boundary value problems and nonlinear partial differential equations based on shooting, finite differences, collocation, and finite elements. These new techniques have only recently begun to be applied to real problems, and have found significant application in solving some engineering analysis problems.

Homotopy methods are very powerful, robust, accurate, numerically stable, and almost universally applicable, but also often prohibitively expensive. They are particularly suitable for highly nonlinear problems for which initial solution estimates are difficult to obtain. Properly implemented they are indeed *globally convergent*, i.e., converge to a solution from an *arbitrary* starting point. This (costly) global convergence feature is their forte, but also makes them inappropriate for mildly nonlinear problems or problems for which a good initial estimate of the solution is easily obtained.

As models of physical phenomena become more ambitious, and supercomputing capability becomes more widespread, it becomes increasingly necessary to use sophisticated numerical analysis tools to solve the mathematical models efficiently. Since nonlinear systems of equations are ubiquitous in engineering analysis, the relevance of homotopy methods is clear. As models become larger and more complicated, and initial solution estimates harder to obtain, the global convergence from an arbitrary starting point property of homotopy algorithms makes them more attractive. Robustness and global convergence are not without cost, however. This extra cost, which weighs against homotopy methods on serial computers, becomes an *advantage* on parallel (and to a lesser

† Department of Computer Science, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061 USA. This work was supported in part by DOE Grant DE-FG05-88ER25068, NASA Grant NAG-1-1079, and AFOSR Grant 89-0497.

extent on vector) computers because there is a large amount of inherent parallelism in homotopy algorithms.

The purpose of this paper is to introduce the theory of globally convergent homotopy methods relevant to engineering analysis, to describe some available computer software, and to give some actual engineering applications. Section 2 gives an intuitive explanation of what is different about the new globally convergent homotopy algorithms, and briefly recounts the basic mathematical theory. Section 3 outlines some numerical algorithms implemented in the mathematical software package HOMPACK. Examples of the globally convergent homotopy techniques applied to realistic engineering problems are presented in some detail in Section 4.

2. Survey of basic homotopy theory.

2.1. Continuation versus homotopy. Continuation is a well known and established procedure in numerical analysis. The idea is to continuously deform a simple (easy) problem into the given (hard) problem, while solving the family of deformed problems. The solutions to the deformed problems are related, and can be tracked as the deformation proceeds. The function describing the deformation is called a *homotopy map*. Homotopies are a traditional part of topology, and have found significant application in nonlinear functional analysis and differential geometry. Similar ideas, such as incremental loading, are also widely used in engineering.

These traditional continuation algorithms have serious deficiencies, which have been removed by modern homotopy algorithms. The differences, however, are subtle and mathematically deep, and the mathematical proofs of the statements in this article are beyond the scope of the presentation here. To explain the differences between the old and new homotopy techniques, a more detailed discussion is required. Suppose the given problem is to find a root of the nonlinear equation $f(x) = 0$, and that $s(x) = 0$ is a simple version of the given problem with an easily obtainable unique solution x_0. Then a homotopy map could be, e.g.,

$$H(\lambda, x) = \lambda f(x) + (1 - \lambda) s(x), \quad 0 \le \lambda \le 1.$$

The family of problems is $H(\lambda, x) = 0$, $0 \le \lambda \le 1$, and the idea would be to track the solutions of $H(\lambda, x) = 0$, starting from $(\lambda, x) = (0, x_0)$, as λ goes from 0 to 1. If everything worked out well, this would lead to a point $(\lambda, x) = (1, \bar{x})$, where $f(\bar{x}) = 0$. The "standard" approach is to start from a point (λ_i, x_i) with $H(\lambda_i, x_i) = 0$, and solve the problem $H(\lambda_i + \Delta\lambda, x) = 0$ for x, with $\Delta\lambda$ being a sufficiently small, fixed, positive number. The bad things that can happen are:

1) The points (λ_i, x_i) may diverge to infinity as $\lambda \to 1$.

2) The problem $H(\lambda_i + \Delta\lambda, x) = 0$ may be singular at its solution, causing numerical instability.

3) There may be no solution of $H(\lambda_i + \Delta\lambda, x) = 0$ near (λ_i, x_i).

The modern approach to homotopy methods is to construct a homotopy map $\rho_a(\lambda, x)$, involving additional parameters in the vector a, such that 1), 2), and 3) never occur or never cause any difficulty. This is the essence of modern probability-one homotopy algorithms, and the details of the construction are given in the next section.

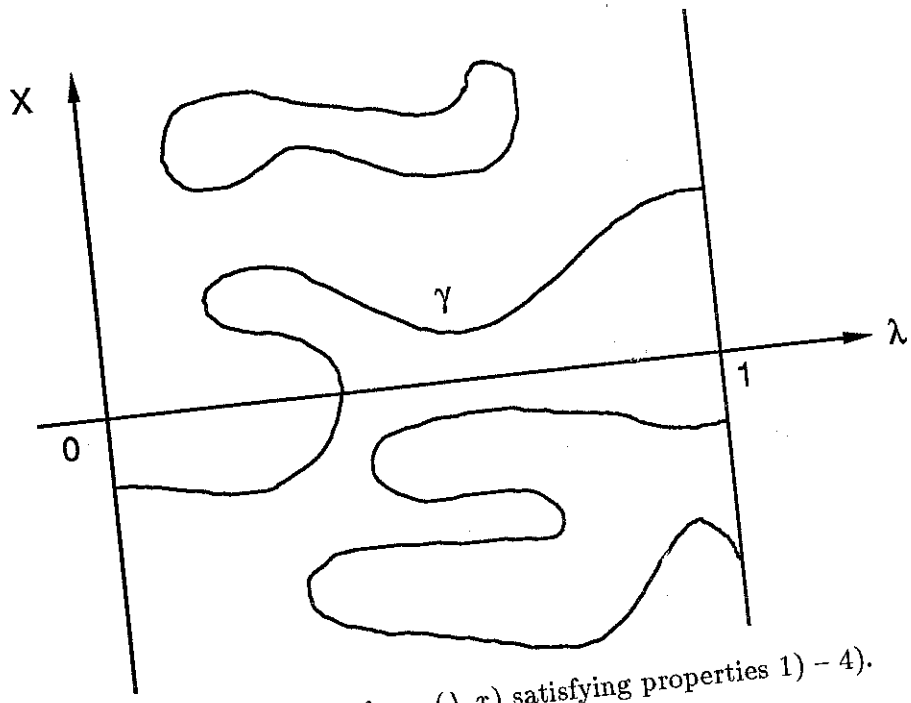


Figure 1. Zero set for $\rho_a(\lambda, x)$ satisfying properties 1) – 4).

2.2. Basic homotopy theorems. The theoretical foundation of all probability one globally convergent homotopy methods is given in the following differential geometry theorem:

DEFINITION. *Let E^n denote n-dimensional real Euclidean space, let $U \subset E^m$ and $V \subset E^n$ be open sets, and let $\rho : U \times [0,1) \times V \to E^n$ be a C^2 map. ρ is said to be transversal to zero if the Jacobian matrix $D\rho$ has full rank on $\rho^{-1}(0)$.*

PARAMETRIZED SARD'S THEOREM [5]. *If $\rho(a, \lambda, x)$ is transversal to zero, then for almost all $a \in U$ the map*

$$\rho_a(\lambda, x) = \rho(a, \lambda, x)$$

is also transversal to zero; i.e., with probability one the Jacobian matrix $D\rho_a(\lambda, x)$ has full rank on $\rho_a^{-1}(0)$.

The import of this theorem is that the zero set $\rho_a^{-1}(0)$ consists of smooth, nonintersecting curves in $[0,1) \times V$. These curves are either closed loops, or have endpoints in $\{0\} \times V$ or $\{1\} \times V$, or go to infinity. Another important consequence is that these curves have finite arc length in any compact subset of $[0,1) \times V$. The recipe for constructing a globally convergent homotopy algorithm to solve the nonlinear system of equations

$$F(x) = 0, \tag{1}$$

where $F : E^n \to E^n$ is a C^2 map, is as follows: For an open set $U \subset E^m$ construct a C^2 homotopy map $\rho : U \times [0,1) \times E^n \to E^n$ such that

1) $\rho(a, \lambda, x)$ is transversal to zero,
2) $\rho_a(0, x) = \rho(a, 0, x) = 0$ is trivial to solve and has a unique solution x_0,
3) $\rho_a(1, x) = F(x)$,
4) $\rho_a^{-1}(0)$ is bounded.

Then for almost all $a \in U$ there exists a zero curve γ of ρ_a, along which the Jacobian matrix $D\rho_a$ has rank n, emanating from $(0, x_0)$ and reaching a zero $\bar{x}$ of F at $\lambda = 1$. This zero curve γ does not intersect itself, is disjoint from any other zeros of ρ_a, and has finite arc length in every compact subset of $[0,1) \times E^n$. Furthermore, if $DF(\bar{x})$ is nonsingular, then γ has finite arc length. See Figure 1.

The general idea of the algorithm is now apparent: just follow the zero curve γ emanating from $(0, x_0)$ until a zero $\tilde{x}$ of $F(x)$ is reached (at $\lambda = 1$). Of course it is nontrivial to develop a viable numerical algorithm based on that idea, but at least conceptually, the algorithm for solving the nonlinear system of equations $F(x) = 0$ is clear and simple. The homotopy map (usually, but not always) is

$$\rho_a(\lambda, x) = \lambda F(x) + (1 - \lambda)(x - a), \tag{2}$$

which has the same form as a standard continuation or embedding mapping. However, there are two crucial differences. First, in standard continuation, the embedding parameter λ increases monotonically from 0 to 1 as the trivial problem $x - a = 0$ is continuously deformed to the problem $F(x) = 0$. The present homotopy method permits λ to both increase and decrease along γ with no adverse effect; that is, turning points present no special difficulty. The second important difference is the use of the extraneous parameter a, whose consequence is that there are never any "singular points" which afflict standard continuation methods. The way in which the zero curve γ of ρ_a is followed and the full rank of $D\rho_a$ along γ guarantee this.

In order for property 4) above to hold for the homotopy map in (2), $F(x)$ and $(x - a)$ must be "asymptotically similar" (see Lemma 3 below). This is not the case for *every* $F(x)$, and so frequently other homotopy maps must be used, for example,

$$\rho_a(\lambda, x) = \lambda F(x) + (1 - \lambda) G(x; a), \tag{2a}$$

where $G(x; a)$ is a simple version of $F(x)$. For instance, $G(x; a)$ might be derived by simplifying the physical model used to derive $F(x)$. Also the homotopy map need not be a simple convex combination between $F(x)$ and $G(x; a)$; examples of homotopy maps nonlinear in λ are in [57] and [64].

The scheme just described is known as a probability-one globally convergent homotopy algorithm. The phrase "probability-one" refers to the almost any choice for a, and the "global convergence" refers to the fact that the starting point x_0 need not be anywhere near the solution $\bar{x}$. It should be emphasized that the form of the homotopy map $\rho_a(\lambda, x)$ in (2) is just a special case used here for clarity of exposition. The more general theory can be found in [66], [70–72], [74], and practical engineering problems requiring a ρ_a nonlinear in λ are in [57] and [64]. Below are some typical theorems for various classes of problems.

The computation of Brouwer fixed points represents one of the first successes for both simplicial [1], [33] and continuous homotopy methods [5], [66]. Brouwer fixed point problems can be very nasty, and often cause locally convergent iterative methods a great deal of difficulty.

THEOREM [5]. *Let $B = \{x \in E^n \mid \|x\|_2 = 1\}$ be the closed unit ball, and $f : B \to B$ a C^2 map. Then for almost all $a \in \text{int } B$ there exists a zero curve γ of*

$$\rho_a(\lambda, x) = \lambda(x - f(x)) + (1 - \lambda)(x - a),$$

along which the Jacobian matrix $D\rho_a(\lambda, x)$ has full rank, emanating from $(0, a)$ and reaching a fixed point $\bar{x}$ of f at $\lambda = 1$. Furthermore, γ has finite arc length if $I - Df(\bar{x})$ is nonsingular.

Typically a mathematical problem (such as a partial differential equation) results in a finite dimensional nonlinear system of equations, and what is desired are conditions on the original problem, not on the final discretized problem. Thus the results in this section are used to derive, working backwards, useful conditions on the original problem, whatever it might be. The following four lemmas, which follow from the results of [5], are used for that purpose.

LEMMA 1. *Let $g : E^p \to E^p$ be a C^2 map, $a \in E^p$, and define $\rho_a : [0,1) \times E^p \to E^p$ by*

$$\rho_a(\lambda, y) = \lambda g(y) + (1-\lambda)(y-a).$$

Then for almost all $a \in E^p$ there is a zero curve γ of ρ_a emanating from $(0,a)$ along which the Jacobian matrix $D\rho_a(\lambda, y)$ has full rank.

LEMMA 2. *If the zero curve γ in Lemma 1 is bounded, it has an accumulation point $(1, \bar{y})$, where $g(\bar{y}) = 0$. Furthermore, if $Dg(\bar{y})$ is nonsingular, then γ has finite arc length.*

LEMMA 3. *Let $F : E^p \to E^p$ be a C^2 map such that for some $r > 0$, $x\,F(x) \geq 0$ whenever $\|x\| = r$. Then F has a zero in $\{x \in E^p \mid \|x\| \leq r\}$, and for almost all $a \in E^p$, $\|a\| < r$, there is a zero curve γ of*

$$\rho_a(\lambda, x) = \lambda F(x) + (1-\lambda)(x-a),$$

along which the Jacobian matrix $D\rho_a(\lambda, x)$ has full rank, emanating from $(0,a)$ and reaching a zero $\bar{x}$ of F at $\lambda = 1$. Furthermore, γ has finite arc length if $DF(\bar{x})$ is nonsingular.

Lemma 3 is a special case of the following more general lemma.

LEMMA 4. *Let $F : E^p \to E^p$ be a C^2 map such that for some $r > 0$ and $\tilde{r} > 0$, $F(x)$ and $x - a$ do not point in opposite directions for $\|x\| = r$, $\|a\| < \tilde{r}$. Then F has a zero in $\{x \in E^p \mid \|x\| \leq r\}$, and for almost all $a \in E^p$, $\|a\| < \tilde{r}$, there is a zero curve γ of*

$$\rho_a(\lambda, x) = \lambda F(x) + (1-\lambda)(x-a),$$

along which the Jacobian matrix $D\rho_a(\lambda, x)$ has full rank, emanating from $(0,a)$ and reaching a zero $\bar{x}$ of F at $\lambda = 1$. Furthermore, γ has finite arc length if $DF(\bar{x})$ is nonsingular.

These theoretical algorithms have been implemented in production quality mathematical software packages such as PITCON [32], CONKUB [22], and HOMPACK [47]. The latter, described in Section 3, is an extensive collection of FORTRAN 77 routines implementing three different tracking algorithms for problems with both dense and sparse Jacobian matrices, and containing high level drivers for special classes of problems.

2.3.1. Two-point boundary value problems: shooting. The next few sections consider nonlinear two-point boundary value problems of the form

$$y''(t) = g\big(t, y(t), y'(t)\big), \quad 0 \leq t \leq 1, \tag{3}$$

$$y(0) = 0, \qquad y(1) = 0 \quad (\text{or } y'(1) = 0), \tag{4}$$

where $y = (y_1, \ldots, y_n)$, $g(t,u,v)$ satisfies a Lipschitz condition in (u,v) for $0 \leq t \leq 1$, and has continuous second partials with respect to u and v for $0 \leq t \leq 1$. These technical assumptions vary slightly from theorem to theorem and method to method; consult the references for complete and precise statements of the theorems. The intent here is to provide the flavor of applying homotopy methods to approximations to nonlinear two-point boundary value problems, and not get bogged down in technical detail.

Let $u(t)$ be the unique solution of

$$y''(t) = g\big(t, y(t), y'(t)\big), \quad 0 \leq t \leq 1,$$

$$y(0) = 0, \quad y'(0) = v,$$

and define $f : E^n \to E^n$ by

$$f(v) = u(1) \quad (\text{or } u'(1)\,). \tag{5}$$

Observe that the original problem (3)–(4) is equivalent to solving

$$f(v) = 0 \tag{6}$$

for the correct initial condition $\bar{v}$.

THEOREM. *Suppose there exists a constant $M > 0$ such that $\|g(t, x(t), x'(t))\|_\infty \leq M$ along every trajectory $x(t)$ for which $x(0) = 0$, $\|x'(0)\|_\infty = M$. Then for almost all $w \in E^n$ with $\|w\|_\infty < M$ there exists a zero curve γ of the homotopy map*

$$\rho_w(\lambda, v) = \lambda f(v) + (1 - \lambda)(v - w),$$

along which $D\rho_w$ has full rank, lying in $[0,1] \times \{v \in E^n \mid \|v\|_\infty \leq M\}$ and connecting $(0, w)$ to $(1, \bar{v})$, where $\bar{v}$ is a zero of f. If $Df(\bar{v})$ is nonsingular, then γ has finite arc length.

2.3.2. Two-point boundary value problems: finite differences. The problem under consideration is (3)–(4). Using standard centered second order accurate finite difference approximations for y' and y'' in (3) with the boundary conditions (4) at mesh points $0 = t_0 < t_1 < \cdots < t_N < t_{N+1} = 1$ results in the finite difference approximation

$$G(Y) = AY + h^2 F^h(Y) = 0, \tag{7}$$

where $h = \max_i(t_{i+1} - t_i)$ is the mesh size, A is a constant positive definite matrix, and $F^h(Y)$ is the nonlinear part of $G(Y)$ due to g.

THEOREM. *Let $F^h(Y)$ be a C^2 mapping, and suppose that*

$$\limsup_{\|Y\|_2 \to \infty} \frac{\|F^h(Y)\|_2}{\|Y\|_2} \leq 9.$$

Then for almost $W \in E^N$ there is a zero curve γ of the homotopy map

$$\rho_W(\lambda, Y) = \lambda G(Y) + (1 - \lambda)(Y - W),$$

along which the Jacobian matrix $D\rho_W(\lambda, Y)$ has full rank, emanating from $(0, W)$ and reaching a zero $\tilde{Y}$ of G at $\lambda = 1$. Furthermore, γ has finite arc length if $DG(\tilde{Y})$ is nonsingular.

More general boundary conditions than (4) have the form

$$By(0) + B'y'(0) + Cy(1) + C'y'(1) = b, \tag{8}$$

where rank $(B\ B'\ C\ C') = 2n$, and there are other technical restrictions (see Keller [13]). These boundary conditions lead to a different nonlinear system of equations

$$G(Y) = AY + h^2 F^h(Y) = 0, \tag{9}$$

where A and F^h are different from those in (7).

THEOREM. *Let $F^h(Y)$ be a C^2 mapping, and suppose that G from (9) satisfies one of the following:*

1) *there exist $r > 0$ and $\tilde{r} > 0$ such that $Y - W$ and $G(Y)$ do not point in opposite directions for $\|Y\|_2 = r$, $\|W\|_2 < \tilde{r}$;*
2) *there exists $r > 0$ such that $Y^t G(Y) \geq 0$ for $\|Y\|_2 = r$;*
3) *A is positive semidefinite, and there exists $r > 0$ such that $Y^t F^h(Y) \geq 0$ for $\|Y\|_2 = r$.*

Then for almost all $\|W\|_2 < \tilde{r}$ there is a zero curve γ of the homotopy map

$$\rho_W(\lambda, Y) = \lambda G(Y) + (1 - \lambda)(Y - W),$$

along which the Jacobian matrix $D\rho_W(\lambda, Y)$ has full rank, emanating from $(0, W)$ and reaching a zero $\tilde{Y}$ of G at $\lambda = 1$. Furthermore, γ has finite arc length if $DG(\tilde{Y})$ is nonsingular.

THEOREM. *The conclusion of the previous theorem holds if A from (9) is positive definite and $g(t, u, v)$ is a bounded C^2 mapping.*

THEOREM. *The conclusion of the previous theorem also holds if $g(t, u, v)$ is a bounded C^2 mapping and the boundary conditions are of the form $y(0) = b_1$, $y(1) = b_2$.*

2.3.3. Two-point boundary value problems: collocation. The idea here is to approximate $y(t)$ from some convenient, finite dimensional vector space S_m with basis $\varphi_1, \ldots, \varphi_m$. These basis functions could be orthogonal polynomials (a popular approach in chemical engineering) or B-splines (the numerical analysts' choice) or something tailored for a specific problem. The approximations

$$y_k(t) \approx A_k(t) = \sum_{i=1}^{m} \alpha_{ki} \varphi_i(t), \quad k = 1, \ldots, n, \tag{10}$$

are substituted into equations (3)–(4) evaluated at discrete points, the collocation points, in the interval $[0, 1]$. This results in the nonlinear system of equations

$$F(Y) = M\,Y + N(Y) = 0, \tag{11}$$

where

$$Y = (\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1m}, \alpha_{21}, \ldots, \alpha_{2m}, \ldots, \alpha_{n1}, \ldots, \alpha_{nm})^t,$$

M is a constant matrix, and $N(Y)$ is the nonlinear part due to g. The dimension of the problem is $p = nm$.

THEOREM. *Let $N(Y)$ in (11) be a C^2 mapping, and suppose there exist constants C and ν such that*

$$\limsup_{\|Y\|_\infty \to \infty} \frac{\|N(Y)\|_\infty}{\|Y\|_\infty^\nu} = C, \qquad 0 \le \nu < 1. \tag{12}$$

For $W \in E^p$, define $\rho_W : [0, 1) \times E^p \to E^p$ by

$$\rho_W(\lambda, Y) = \lambda\, F(Y) + (1 - \lambda)(Y - W).$$

Then for almost all $W \in E^p$ there exists a zero curve γ of ρ_W, along which the Jacobian matrix $D\rho_W(\lambda, Y)$ has full rank, emanating from $(0, W)$ and reaching a zero $\tilde{Y}$ of F (at $\lambda = 1$). Furthermore, if $DF(\tilde{Y})$ is nonsingular, then γ has finite arc length.

THEOREM. *Let $N(Y)$ in (11) be a C^2 mapping and the matrix M be such that*

$$\min_{\|Y\|_\infty = 1} \max_{1 \le j \le p} Y_j\,(MY)_j = \Gamma > 0.$$

If

$$\limsup_{\|Y\|_\infty \to \infty} \frac{\|N(Y)\|_\infty}{\|Y\|_\infty} = C < \Gamma,$$

then the conclusion of the above Theorem holds.

THEOREM. *If $g(t, u, v)$ in (3) is C^2 and bounded, then the conclusion of the above Theorem holds.*

THEOREM. *Let $g(t, u, v)$ in (3) be a C^2 mapping, and suppose there exist constants μ and ν such that*

$$\limsup_{\|y\|_\infty \to \infty} \max_{\substack{0 \le t \le 1 \\ u \in E^n}} \frac{\|g(t, y, u)\|_\infty}{\|y\|_\infty^\nu} = \mu, \quad 0 \le \nu < 1. \tag{13}$$

Then the conclusion of the above Theorem holds.

2.3.4. Two-point boundary value problems: finite elements. The finite element (or Rayleigh-Ritz-Galerkin) approach is similar to collocation in that an approximation is sought from a finite dimensional space S_m. Here the elements of S_m automatically satisfy the boundary conditions (4), which collocation doesn't require. Instead of satisfying the differential equation (3) at discrete points, the finite element method satisfies (3) in an average sense by requiring certain inner product integrals to be equal. In some contexts the finite element formulation can be viewed as minimizing some functional over a finite dimensional space, where the minimum over an infinite dimensional space gives the exact solution (a variational formulation).

Using the approximations (10), the Galerkin approximation to (3)–(4) is

$$\int_0^1 A_j''(x)\varphi_k(x)\,dx = \int_0^1 g_j\big(x, A(x), A'(x)\big)\varphi_k(x)\,dx, \qquad k = 1,\ldots,m, \quad j = 1,\ldots,n. \tag{14}$$

This is a system of equations of exactly the same form as (11). The convergence theorems are similar to those for collocation.

THEOREM. *Let $N(Y)$ in (11) be a C^2 mapping, and suppose there exist constants C and ν such that*

$$\limsup_{\|Y\|_2\to\infty} \frac{\|N(Y)\|_2}{\|Y\|_2^{\nu}} = C, \qquad 0 \le \nu < 1.$$

For $W \in E^p$, define $\rho_W : [0,1) \times E^p \to E^p$ by

$$\rho_W(\lambda, Y) = \lambda F(Y) + (1-\lambda)(Y - W).$$

Then for almost all $W \in E^p$ there exists a zero curve γ of ρ_W, along which the Jacobian matrix $D\rho_W(\lambda, Y)$ has full rank, emanating from $(0, W)$ and reaching a zero $\tilde{Y}$ of F (at $\lambda = 1$). Furthermore, if $DF(\tilde{Y})$ is nonsingular, then γ has finite arc length.

THEOREM. *Let $N(Y)$ in (11) be a C^2 mapping and $\Gamma > 0$ the smallest eigenvalue of M. If*

$$\limsup_{\|Y\|_2\to\infty} \frac{\|N(Y)\|_2}{\|Y\|_2} = C < \Gamma,$$

then the conclusion of the above Theorem holds.

THEOREM. *If $g(t,u,v)$ in (3) is C^2 and bounded, then the conclusion of the above Theorem holds.*

THEOREM. *Let $g(t,u,v)$ in (3) be a C^2 mapping, and suppose there exist constants μ, ξ, and ν such that*

$$\|g(t,u,v)\|_2 \le \mu\big(\xi + (\|u\|_2 + \|v\|_2)^{\nu}\big), \quad 0 \le \nu < 1, \tag{15}$$

for all $0 \le t \le 1$ and $u, v \in E^n$. Then the conclusion of the above Theorem holds.

2.4. Basic optimization homotopies. Consider first the unconstrained optimization problem

$$\min_x f(x). \tag{16}$$

THEOREM [68]. *Let $f : E^n \to E$ be a C^3 convex map with a minimum at $\tilde{x}$, $\|\tilde{x}\|_2 \leq M$. Then for almost all a, $\|a\|_2 < M$, there exists a zero curve γ of the homotopy map*

$$\rho_a(\lambda, x) = \lambda \nabla f(x) + (1-\lambda)(x-a),$$

along which the Jacobian matrix $D\rho_a(\lambda, x)$ has full rank, emanating from $(0, a)$ and reaching a point $(1, \tilde{x})$, where $\tilde{x}$ solves (16).

A function is called uniformly convex if it is convex and its Hessian's smallest eigenvalue is bounded away from zero. Consider next the constrained optimization problem

$$\min_{x \geq 0} f(x). \tag{17}$$

This is more general than it might appear because the general convex quadratic program reduces to a problem of the form (17).

THEOREM [68]. *Let $f : E^n \to E$ be a C^3 uniformly convex map. Then there exists $\delta > 0$ such that for almost all $a \geq 0$ with $\|a\|_2 < \delta$ there exists a zero curve γ of the homotopy map*

$$\rho_a(\lambda, x) = \lambda K(x) + (1-\lambda)(x-a),$$

where

$$K_i(x) = -\left|\frac{\partial f(x)}{\partial x_i} - x_i\right|^3 + \left(\frac{\partial f(x)}{\partial x_i}\right)^3 + x_i^3,$$

along which the Jacobian matrix $D\rho_a(\lambda, x)$ has full rank, connecting $(0, a)$ to a point $(1, \bar{x})$, where $\bar{x}$ solves the constrained optimization problem (17).

Given $F : E^n \to E^n$, the nonlinear complementarity problem is to find a vector $x \in E^n$ such that

$$x \geq 0, \quad F(x) \geq 0, \quad x^t F(x) = 0. \tag{18}$$

At a solution $\bar{x}$, $\bar{x}$ and $F(\bar{x})$ are "complementary" in the sense that if $\bar{x}_i > 0$, then $F_i(\bar{x}) = 0$, and if $F_i(\bar{x}) > 0$, then $\bar{x}_i = 0$. This problem is difficult because there are linear constraints $x \geq 0$, nonlinear constraints $F(x) \geq 0$, and a combinatorial aspect from the complementarity condition $x^t F(x) = 0$. It is interesting that homotopy methods can be adapted to deal with nonlinear constraints and combinatorial conditions.

Define $G : E^n \to E^n$ by

$$G_i(z) = -\left|F_i(z) - z_i\right|^3 + \left(F_i(z)\right)^3 + z_i^3, \quad i = 1, \ldots, n,$$

and let

$$\rho_a(\lambda, z) = \lambda G(z) + (1-\lambda)(z-a).$$

THEOREM [70]. *Let $F : E^n \to E^n$ be a C^2 map, and let the Jacobian matrix $DG(z)$ be nonsingular at every zero of $G(z)$. Suppose there exists $r > 0$ such that $z > 0$ and $z_k = \|z\|_\infty \geq r$ imply $F_k(z) > 0$. Then for almost all $a > 0$ there exists a zero curve γ of $\rho_a(\lambda, z)$, along which the Jacobian matrix $D\rho_a(\lambda, z)$ has full rank, having finite arc length and connecting $(0, a)$ to $(1, \bar{z})$, where $\bar{z}$ solves (18).*

THEOREM [70]. *Let $F : E^n \to E^n$ be a C^2 map, and let the Jacobian matrix $DG(z)$ be nonsingular at every zero of $G(z)$. Suppose there exists $r > 0$ such that $z \geq 0$ and $\|z\|_\infty \geq r$ imply $z_k F_k(z) > 0$ for some index k. Then there exists $\delta > 0$ such that for almost all $a \geq 0$ with $\|a\|_\infty < \delta$ there exists a zero curve γ of $\rho_a(\lambda, z)$, along which the Jacobian matrix $D\rho_a(\lambda, z)$ has full rank, having finite arc length and connecting $(0, a)$ to $(1, \bar{z})$, where $\bar{z}$ solves (18).*

Homotopy algorithms for convex unconstrained optimization are only of theoretical interest, and are generally not computationally competitive with other approaches, but it is reassuring that the globally convergent homotopy techniques can theoretically be directly applied. For constrained optimization the homotopy approach offers some advantages, and, especially for the nonlinear complementarity problem, is competitive with other algorithms. See [48] for an application of homotopy techniques to the linear complementarity problem. Constrained optimization is addressed in the next few sections.

2.5. Expanded Lagrangian Homotopy. The expanded Lagrangian homotopy method of Poore [30], [31] is applicable to the general nonlinear programming problem

$$\begin{aligned} &\min \theta(x) \\ \text{subject to} \quad &g(x) \leq 0, \\ &h(x) = 0, \end{aligned}$$

where $x \in E^n$, θ is real valued, g is an m-dimensional vector, and h is a p-dimensional vector. Assume that θ, g, and h are C^2. In this general situation the complete formulation and solution algorithm for the expanded Lagrangian homotopy are rather complicated. The essence of the method is presented here, referring the reader to [30] and [31] for a discussion of the theoretical and practical subtleties. The technique has been applied to linear programming [30] and the linear complementarity problem [48], but is currently primarily of theoretical interest.

The expanded Lagrangian approach may be described as an optimization/continuation approach and has in its simplest form two main steps.

Step 1. (Optimization phase).

At $r = r_0 > 0$ solve the unconstrained minimization problem

$$\min_x P(x, r)$$

where

$$P(x, r) = \theta(x) + \frac{1}{2r} h(x)^t h(x) - r \sum_{i=1}^{m} \ln\bigl(-g_i(x)\bigr).$$

Step 2A. (Switch to expanded system).

A (local) solution of $\min P$ must satisfy

$$0 = \nabla_x P = \nabla\theta(x) + \frac{h(x)^t \nabla h(x)}{r} - \sum_{i=1}^{m} \frac{r}{g_i(x)} \nabla g_i(x).$$

Introduce the following variables:

$$\beta = \frac{h(x)}{r},$$
$$\mu_i = \frac{r}{-g_i(x)}, \qquad i = 1, \ldots, m,$$

which ultimately represent the Lagrange multipliers. This helps to remove the inevitable ill-conditioning associated with penalty methods for small r and we thus obtain our equivalent but expanded system:

$$\begin{aligned} \nabla\theta(x) + \beta^t \nabla h(x) + \mu^t \nabla g(x) &= 0, \\ h(x) - r\beta &= 0, \\ \mu_i g_i(x) + r &= 0, \qquad i = 1, \ldots, m. \end{aligned}$$

(Remark. As a result of the optimization phase and the initial starting point with $r_0 > 0$, the solution $x^{(0)}$ of $\min P(x, r_0)$ satisfies $g(x^{(0)}) < 0$. As a consequence, $\mu^{(0)} > 0$ from the definition of μ. μ remains positive until $r = 0$ where we formally have

$$\begin{aligned} \nabla\theta(x) + \beta^t \nabla h(x) + \mu^t \nabla g(x) &= 0, \\ h(x) &= 0, \\ g(x) &\leq 0, \\ \mu &\geq 0, \\ \mu_i g_i(x) &= 0, \qquad i = 1, \ldots, m, \end{aligned}$$

which implies that we have solved the problem.)

In practice we do not solve the optimization problem $\min P$ to high accuracy since a highly accurate solution may have only a digit or two in common with the final answer. However, it is imperative that ∇P be reasonably small in magnitude, say less than $r_0/10$. The expanded system is converted to a homotopy map by letting $r = r_0(1-\lambda)$ and modifying the first equation to obtain:

$$\begin{aligned} \nabla\theta(x) + \beta^t \nabla h(x) + \mu^t \nabla g(x) - \frac{r}{r_0} \nabla P(x^{(0)}, r_0) &= 0, \\ h(x) - r\beta &= 0, \\ \mu_i g_i(x) + r &= 0, \qquad i = 1, \ldots, m. \end{aligned} \tag{19}$$

Write this system of $n + p + m$ equations in the $n + p + m + 1$ variables λ, x, β, μ as

$$\Upsilon(\lambda, x, \beta, \mu) = 0.$$

Step 2B. (Track the zero curve of Υ from $r = r_0$ to $r = 0$.)

Starting with arbitrary $r_0 > 0$ and feasible interior point $x^{(0)}$ $(g(x^{(0)}) < 0)$, the rest of the initial point $(0, x^{(0)}, \beta^{(0)}, \mu^{(0)})$ is given by

$$\beta^{(0)} = \frac{h(x^{(0)})}{r_0},$$
$$\mu_i^{(0)} = \frac{r_0}{-g_i(x^{(0)})}, \qquad i = 1, \ldots, m.$$

This approach requires careful attention to implementation details. For example, the linear algebra and globalization techniques with dynamic scaling are critically important in the optimization phase. For degenerate problems the path can still be long. One possible resolution is the use of shifts and weights as developed in the method of multipliers [3], but holding $r = r_0$ fixed. (This approach is currently under investigation in the context of linear programming [30].) Note that the optimization phase (Step 1) can be omitted altogether, starting Step 2B with an arbitrary interior feasible point $x^{(0)}$ $(g(x^{(0)}) < 0)$, so that (19) is a true global homotopy. As a practical matter, however, it is advantageous to get a good starting point by doing Step 1 with a small r_0.

2.5.1. Application of expanded Lagrangian homotopy to the linear complementarity problem. As an illustration, the expanded Lagrangian homotopy method will be applied to the linear complementarity problem:

$$w - Mz = q,$$
$$w \geq 0, \quad z \geq 0, \quad w^t z = 0,$$

where M is a given real $n \times n$ matrix and $q \in E^n$ is given; the unknowns are $w \in E^n$ and $z \in E^n$.

Step 1. (Optimization phase).

At $r = r_0 > 0$ solve the unconstrained minimization problem

$$\min_{w,z} P(w, z, r)$$

where

$$P(w, z, r) = \frac{1}{2r}\|w - Mz - q\|_2^2 + \frac{1}{2r}\langle w, z\rangle^2 - r\sum_{i=1}^{n} \ln z_i - r\sum_{i=1}^{n} \ln w_i.$$

Step 2A. (Switch to expanded system).

A (local) solution of $\min P$ must satisfy

$$0 = \nabla_{(w,z)} P = \begin{pmatrix} I \\ -M^t \end{pmatrix} \frac{(w - Mz - q)}{r} + \begin{pmatrix} z \\ w \end{pmatrix} \frac{\langle w, z\rangle}{r} - r\left(\frac{1}{w_1}, \ldots, \frac{1}{w_n}, \frac{1}{z_1}, \ldots, \frac{1}{z_n}\right)^t.$$

Introduce the following variables:

$$\beta = \frac{w - Mz - q}{r},$$
$$\theta = \frac{\langle w, z\rangle}{r},$$
$$\mu_i = \frac{r}{w_i}, \qquad i = 1, \ldots, n,$$
$$\eta_i = \frac{r}{z_i}, \qquad i = 1, \ldots, n,$$

which ultimately represent the Lagrange multipliers. This helps to remove the inevitable ill-conditioning associated with penalty methods for small r and we thus obtain our equivalent but expanded system:

$$\begin{aligned}
\begin{pmatrix} I \\ -M^t \end{pmatrix} \beta + \begin{pmatrix} z \\ w \end{pmatrix} \theta - \begin{pmatrix} \mu \\ \eta \end{pmatrix} &= 0, \\
w - Mz - q - r\beta &= 0, \\
\langle w, z \rangle - r\theta &= 0, \\
\mu_i w_i - r &= 0, \qquad i = 1, \ldots, n, \\
\eta_i z_i - r &= 0, \qquad i = 1, \ldots, n.
\end{aligned}$$

(Remark. As a result of the optimization phase and the initial starting point with $r_0 > 0$, the solution $(w^{(0)}, z^{(0)})$ of $\min P(w, z, r_0)$ satisfies $z^{(0)} > 0$ and $w^{(0)} > 0$. As a consequence, $\mu^{(0)} > 0$ and $\eta^{(0)} > 0$ from the definitions of μ and η. They remain positive until $r = 0$ where we formally have

$$\begin{aligned}
\begin{pmatrix} I \\ -M^t \end{pmatrix} \beta + \begin{pmatrix} z \\ w \end{pmatrix} \theta - \begin{pmatrix} \mu \\ \eta \end{pmatrix} &= 0, \\
w - Mz - q &= 0, \\
\langle w, z \rangle &= 0, \\
\mu_i w_i &= 0, \qquad i = 1, \ldots, n, \\
\eta_i z_i &= 0, \qquad i = 1, \ldots, n, \\
w, z, \theta, \mu, \eta &\geqq 0,
\end{aligned}$$

which implies that we have solved the problem.)

The expanded system is converted to a homotopy map by letting $r = r_0(1 - \lambda)$ and modifying the first equation to obtain:

$$\begin{aligned}
\begin{pmatrix} I \\ -M^t \end{pmatrix} \beta + \begin{pmatrix} z \\ w \end{pmatrix} \theta - \begin{pmatrix} \mu \\ \eta \end{pmatrix} - \frac{r}{r_0} \nabla P(w^{(0)}, z^{(0)}, r_0) &= 0, \\
w - Mz - q - r\beta &= 0, \\
\langle w, z \rangle - r\theta &= 0, \\
\mu_i w_i - r &= 0, \qquad i = 1, \ldots, n, \\
\eta_i z_i - r &= 0, \qquad i = 1, \ldots, n.
\end{aligned}$$

Write this system of $5n + 1$ equations in the $5n + 2$ variables λ, w, z, β, θ, μ, η as

$$\Upsilon(\lambda, w, z, \beta, \theta, \mu, \eta) = 0.$$

Step 2B. (Track the zero curve of Υ from $r = r_0$ to $r = 0$.)

Starting with arbitrary $r_0 > 0$, $w^{(0)} > 0$ and $z^{(0)} > 0$, the rest of the initial point $(0, w^{(0)}, z^{(0)}, \beta^{(0)}, \theta_0, \mu^{(0)}, \eta^{(0)})$ is given by

$$\beta^{(0)} = \frac{w^{(0)} - Mz^{(0)} - q}{r_0},$$

$$\theta_0 = \frac{\langle w^{(0)}, z^{(0)} \rangle}{r_0},$$

$$\mu_i^{(0)} = \frac{r_0}{w_i^{(0)}}, \qquad i = 1, \ldots, n,$$

$$\eta_i^{(0)} = \frac{r_0}{z_i^{(0)}}, \qquad i = 1, \ldots, n.$$

Computational experience with this approach to the LCP is reported in [48].

2.6. Kreisselmeier-Steinhauser envelope function. Sections 2.5 and 4.1 present ways that are both theoretically "correct" and computationally "practical" to deal with inequality constraints. However, there are numerous practical difficulties in those approaches, and the implementation and tuning details become absolutely crucial. For example, with the expanded Lagrangian formulation, line searches may generate negative arguments for the ln functions, and the homotopy zero curve may diverge if the Step 1 solution is not good enough. For the active set approach in Section 4.1, the detection and switching criteria for transition points may become extremely cumbersome and inefficient. This section suggests an alternate way of dealing with inequality constraints.

Consider inequality constraints of the form

$$g_i(x) \leq 0, \quad i = 1, \ldots, m, \tag{20}$$

where each $g_i : E^n \to E$ is C^2. For a constant $\rho > 0$, the Kreisselmeier-Steinhauser [14] envelope function for (20) is

$$K(x) = \frac{1}{\rho} \ln \left[\sum_{i=1}^{m} \exp\big(\rho g_i(x)\big) \right]. \tag{21}$$

$K(x)$ is a cumulative measure of the satisfaction or violation of the constraints (20). Let $g_{max}(x) = \max\{g_1(x), \ldots, g_m(x)\}$, and observe that

$$K(x) = g_{max}(x) + \frac{1}{\rho} \ln \left[\sum_{i=1}^{m} \exp\Big(\rho\big(g_i(x) - g_{max}(x)\big)\Big) \right], \tag{22}$$

from which it directly follows that

$$g_{max}(x) \leq K(x) \leq g_{max}(x) + \frac{1}{\rho} \ln m. \tag{23}$$

Thus the envelope $K(x)$ follows the maximum constraint, more closely for large ρ. In particular, (20) could be replaced by

$$K(x) \leq 0 \tag{24}$$

with an error of no more than $(\ln m)/\rho$.

The choice of ρ involves a tradeoff between modelling the maximum constraint (large ρ preferred) and avoiding large gradients (small ρ preferred). If the practical criterion for an active constraint is $|g_i| \leq \epsilon$, then a choice for ρ which has worked well in practice is

$$\rho = \frac{\ln m}{\epsilon}. \tag{25}$$

Observe that $K(x)$ is C^2 and defined *everywhere*, a decided advantage over barrier functions. Furthermore, (24) is a *single* nonlinear constraint, which makes any active set strategy very simple. (24) has been successfully used in large scale structural optimization [2] and optimal control [14].

2.7. Probability-one homotopy for Kuhn-Tucker optimality conditions. The approaches of earlier sections are still not always entirely adequate. The cumulative constraint function (21) is decidedly unnatural, extremely nonlinear and ill conditioned for large ρ, and does not take advantage of a known solution to a related problem. Consider again the general nonlinear programming problem:

$$\begin{aligned} &\min \theta(x) \\ \text{subject to} \quad &g(x) \leq 0, \\ &h(x) = 0, \end{aligned} \tag{26}$$

under the same assumptions mentioned before. The Kuhn-Tucker necessary optimality conditions for (26) are

$$\begin{aligned} \nabla\theta(x) + \beta^t \nabla h(x) + \mu^t \nabla g(x) &= 0, \\ h(x) &= 0, \\ g(x) &\leq 0, \\ \mu &\geq 0, \\ \mu^t g(x) &= 0, \end{aligned} \tag{27}$$

where $\beta \in E^p$ and $\mu \in E^m$. Following Mangasarian [21] and Watson [70], the complementarity conditions $\mu \geq 0$, $g(x) \leq 0$, $\mu^t g(x) = 0$ are replaced by the equivalent nonlinear system of equations

$$W(x, \mu) = 0, \tag{28a}$$

where

$$W_i(x, \mu) = -|\mu_i + g_i(x)|^3 + \mu_i^3 - (g_i(x))^3, \quad i = 1, \ldots, m. \tag{28b}$$

Thus the optimality conditions (27) take the form

$$F(x, \beta, \mu) = \begin{pmatrix} [\nabla\theta(x) + \beta^t \nabla h(x) + \mu^t \nabla g(x)]^t \\ h(x) \\ W(x, \mu) \end{pmatrix} = 0. \tag{29}$$

With $z = (x, \beta, \mu)$, the proposed homotopy map is

$$\rho_a(\lambda, z) = \lambda F(z) + (1 - \lambda)(z - a), \tag{30}$$

where $a \in E^{n+p+m}$. Simple conditions on θ, g, and h guaranteeing that the above homotopy map $\rho_a(\lambda, z)$ will work are unknown, although this map has worked very well on some difficult fuel optimal orbital rendezvous problems [37].

Frequently in practice the functions θ, g, and h involve a parameter vector c, and a solution to (26) is known for some $c = c^{(0)}$. Suppose that the problem under consideration has parameter vector $c = c^{(1)}$. Then

$$c = (1 - \lambda)c^{(0)} + \lambda c^{(1)} \tag{31}$$

parametrizes c by λ and $\theta = \theta(x; c) = \theta(x; c(\lambda))$, $g = g(x; c(\lambda))$, $h = h(x; c(\lambda))$. The optimality conditions in (29) become functions of λ as well, $F(\lambda, x, \beta, \mu) = 0$, and

$$\rho_a(\lambda, z) = \lambda F(\lambda, z) + (1 - \lambda)(z - a) \tag{32}$$

is a highly implicit nonlinear function of λ. If $F(0, z^{(0)}) = 0$, a good choice for a in practice has been found to be $a = z^{(0)}$. A natural choice for a homotopy would be simply

$$F(\lambda, z) = 0, \tag{33}$$

since the solution $z^{(0)}$ to $F(0, z) = 0$ (the problem corresponding to $c = c^{(0)}$) is known. However, for various technical reasons, (32) is much better than (33) [37].

3. Curve tracking algorithms and HOMPACK.

The zero curve γ of the homotopy map $\rho_a(\lambda, x)$ (of which (2) is a special case) can be tracked by many different techniques; refer to the excellent survey [1] and recent work [74], [75]. There are three primary algorithmic approaches to tracking γ that have been used in HOMPACK [47], a software package developed at Sandia National Laboratories, General Motors Research Laboratories, Virginia Polytechnic Institute and State University, and The University of Michigan: 1) an ODE-based algorithm, 2) a predictor-corrector algorithm whose corrector follows the flow normal to the Davidenko flow (a "normal flow" algorithm); 3) a version of Rheinboldt's linear predictor, quasi-Newton corrector algorithm [4], [32], (an "augmented Jacobian matrix" method).

3.1. Ordinary differential equation-based algorithm. Assuming that $F(x)$ is C^2 and a is such that ρ_a is transversal to zero, the zero curve γ is C^1 and can be parametrized by arc length s. Thus $\lambda = \lambda(s)$, $x = x(s)$ along γ, and

$$\rho_a(\lambda(s), x(s)) = 0$$

identically in s. Therefore

$$\frac{d}{ds}\rho_a(\lambda(s), x(s)) = D\rho_a(\lambda(s), x(s)) \begin{pmatrix} \frac{d\lambda}{ds} \\ \frac{dx}{ds} \end{pmatrix} = 0, \tag{34}$$

$$\left\| \left(\frac{d\lambda}{ds}, \frac{dx}{ds} \right) \right\|_2 = 1. \tag{35}$$

With the initial conditions

$$\lambda(0) = 0, \quad x(0) = x_0, \tag{36}$$

the zero curve γ is the trajectory of the initial value problem (34)–(36). When $\lambda(\bar{s}) = 1$, the corresponding $x(\bar{s})$ is a zero of $F(x)$. Thus all the sophisticated ODE techniques currently available can be brought to bear on the problem of tracking γ [34], [66].

Typical ODE software requires $(d\lambda/ds, dx/ds)$ explicitly, and (34), (35) only implicitly define the derivative $(d\lambda/ds, dx/ds)$. Since the dimension of the kernel of the Jacobian matrix

$$D\rho_a(\lambda(s), x(s))$$

is one (this follows from the fact that $D\rho_a$ has full rank p by the Parametrized Sard's Theorem), the derivative $(d\lambda/ds, dx/ds)$ can be calculated from any nonzero vector $z \in \ker D\rho_a$. Note that the derivative $(d\lambda/ds, dx/ds)$ is a unit tangent vector to the zero curve γ. For computational efficiency it is imperative that the number of derivative evaluations be kept small. Complete details for solving the initial value problem (34)–(36) and obtaining $x(\bar{s})$ are given in [49] and [66].

A discussion of the kernel computation follows.

The Jacobian matrix $D\rho_a$ is $p \times (p+1)$ with (theoretical) rank p. The crucial observation is that the last p columns of $D\rho_a$, corresponding to $D_x\rho_a$, may not have rank p, and even if they do, some other p columns may be better conditioned. The objective is to avoid choosing p "distinguished" columns, rather to treat all columns the same (not possible for sparse matrices). There are kernel finding algorithms based on Gaussian elimination and p distinguished columns [15]. Choosing and switching these p columns is tricky, and based on *ad hoc* parameters. Also, computational experience has shown that accurate tangent vectors $(d\lambda/ds, dx/ds)$ are essential, and the accuracy of Gaussian elimination may not be good enough. A conceptually elegant, as well as accurate, algorithm is to compute the QR factorization with column interchanges [74] of $D\rho_a$,

$$Q\, D\rho_a\, P^t P z = \begin{pmatrix} * & \cdots & * & * \\ & \ddots & \vdots & \vdots \\ 0 & & * & * \end{pmatrix} Pz = 0,$$

where Q is a product of Householder reflections and P is a permutation matrix, and then obtain a vector $z \in \ker D\rho_a$ by back substitution. Setting $(Pz)_{p+1} = 1$ is a convenient choice. This scheme provides high accuracy, numerical stability, and a uniform treatment of all $p+1$ columns. Finally,

$$\left(\frac{d\lambda}{ds}, \frac{dx}{ds}\right) = \pm \frac{z}{\|z\|_2},$$

where the sign is chosen to maintain an acute angle with the previous tangent vector on γ. There is a rigorous mathematical criterion, based on a $(p+1) \times (p+1)$ determinant, for choosing the sign, but there is no reason to believe that would be more robust than the angle criterion.

Several features which are a combination of common sense and computational experience should be incorporated into the algorithm. Since most ordinary differential equation solvers only control the local error, the longer the arc length of the zero curve γ gets, the farther away the computed points may be from the true curve γ. Therefore when the arc length gets too long, the last computed point $(\bar{\lambda}, \bar{x})$ is used to calculate a new parameter vector $\bar{a}$ such that

$$\rho_{\bar{a}}(\bar{\lambda}, \bar{x}) = 0 \tag{37}$$

exactly, and the zero curve of $\rho_{\bar{a}}(\lambda, x)$ is followed starting from $(\bar{\lambda}, \bar{x})$. A rigorous justification for this strategy was given in [66]. If ρ_a has the special form in (2), then trivially

$$\bar{a} = \big(\bar{\lambda}\, F(\bar{x}) + (1-\bar{\lambda})\,\bar{x}\big)/(1-\bar{\lambda}).$$

For more general homotopy maps ρ_a, this computation of $\bar{a}$ may be complicated.

Remember that tracking γ was merely a means to an end, namely a zero $\tilde{x}$ of $F(x)$. Since γ itself is of no interest (usually), one should not waste computational effort following it too closely. However, since γ is the only sure way to $\tilde{x}$, losing γ can be disastrous. The tradeoff between computational efficiency and reliability is very delicate, and a fool-proof strategy appears difficult to achieve. None of the three primary algorithms alone is superior overall, and each of the three beats the other two (sometimes by an order of magnitude) on particular problems. Since the algorithms' philosophies are significantly different, a hybrid will be hard to develop.

3.2. Normal flow algorithm. As the homotopy parameter vector a varies, the corresponding homotopy zero curve γ also varies. This family of zero curves is known as the Davidenko flow. The normal flow algorithm is so called because the iterates converge to the zero curve γ along the flow normal to the Davidenko flow (in an asymptotic sense).

The normal flow algorithm has four phases: prediction, correction, step size estimation, and computation of the solution at $\lambda = 1$. For the prediction phase, assume that several points $P^{(1)} = (\lambda(s_1), x(s_1))$, $P^{(2)} = (\lambda(s_2), x(s_2))$ on γ with corresponding tangent vectors $(d\lambda/ds(s_1),$ $dx/ds(s_1))$, $(d\lambda/ds(s_2),\ dx/ds(s_2))$ have been found, and h is an estimate of the optimal step (in arc length) to take along γ. The prediction of the next point on γ is

$$Z^{(0)} = p(s_2 + h), \tag{38}$$

where $p(s)$ is the Hermite cubic interpolating $(\lambda(s), x(s))$ at s_1 and s_2. Precisely,

$$\begin{aligned} p(s_1) &= (\lambda(s_1), x(s_1)), \quad & p'(s_1) &= (d\lambda/ds(s_1), dx/ds(s_1)), \\ p(s_2) &= (\lambda(s_2), x(s_2)), \quad & p'(s_2) &= (d\lambda/ds(s_2), dx/ds(s_2)), \end{aligned}$$

and each component of $p(s)$ is a polynomial in s of degree less than or equal to 3.

Starting at the predicted point $Z^{(0)}$, the corrector iteration is

$$Z^{(k+1)} = Z^{(k)} - \Big[D\rho_a\big(Z^{(k)}\big)\Big]^{\dagger} \rho_a\big(Z^{(k)}\big), \qquad k = 0, 1, \ldots \tag{39}$$

where $\big[D\rho_a(Z^{(k)})\big]^{\dagger}$ is the Moore-Penrose pseudoinverse of the $n \times (n+1)$ Jacobian matrix $D\rho_a$. Small perturbations of a produce small changes in the trajectory γ , and the family of trajectories γ for varying a is known as the "Davidenko flow". Geometrically, the iterates given by (39) return to the zero curve along the flow normal to the Davidenko flow, hence the name "normal flow algorithm".

A corrector step ΔZ is the unique minimum norm solution of the equation

$$\big[D\rho_a\big]\Delta Z = -\rho_a. \tag{40}$$

Fortunately ΔZ can be calculated at the same time as the kernel of $\big[D\rho_a\big]$, and with just a little more work. Normally for dense problems the kernel of $\big[D\rho_a\big]$ is found by computing a QR

factorization of $[D\rho_a]$, and then using back substitution. By applying this QR factorization to $-\rho_a$ and using back substitution again, a *particular* solution v to (40) can be found. Let $u \neq 0$ be any vector in the kernel of $[D\rho_a]$. Then the minimum norm solution of (40) is

$$\Delta Z = v - \frac{v^t u}{u^t u} u. \tag{41}$$

Since the kernel of $[D\rho_a]$ is needed anyway for the tangent vectors, solving (40) only requires another $\mathcal{O}(n^2)$ operations beyond those for the kernel. The number of iterations required for convergence of (39) should be kept small (say < 4) since QR factorizations of $[D\rho_a]$ are expensive. The alternative of using $[D\rho_a(Z^{(0)})]$ for several iterations, which results in linear convergence, is rarely cost effective.

When the iteration (39) converges, the final iterate $Z^{(k+1)}$ is accepted as the next point on γ, and the tangent vector to the integral curve through $Z^{(k)}$ is used for the tangent–this saves a Jacobian matrix evaluation and factorization at $Z^{(k+1)}$. The step size estimation described next attempts to balance progress along γ with the effort expended on the iteration (39).

Define a contraction factor

$$L = \frac{\left\| Z^{(2)} - Z^{(1)} \right\|}{\left\| Z^{(1)} - Z^{(0)} \right\|}, \tag{42}$$

a residual factor

$$R = \frac{\left\| \rho_a(Z^{(1)}) \right\|}{\left\| \rho_a(Z^{(0)}) \right\|}, \tag{43}$$

a distance factor ($Z^* = \lim_{k \to \infty} Z^{(k)}$)

$$D = \frac{\left\| Z^{(1)} - Z^* \right\|}{\left\| Z^{(0)} - Z^* \right\|}, \tag{44}$$

and ideal values $\bar{L}$, $\bar{R}$, $\bar{D}$ for these three. Let h be the current step size (the distance from Z^* to the previous point found on γ), and $\bar{h}$ the "optimal" step size for the next step. The goal is to achieve

$$\frac{\bar{L}}{L} \approx \frac{\bar{R}}{R} \approx \frac{\bar{D}}{D} \approx \frac{\bar{h}^q}{h^q} \tag{45}$$

for some q. This leads to the choice

$$\hat{h} = \left(\min\{\bar{L}/L, \bar{R}/R, \bar{D}/D\} \right)^{1/q} h, \tag{46}$$

a worst case choice. To prevent chattering and unreasonable values, constants $h_{\min}$ (minimum allowed step size), $h_{\max}$ (maximum allowed step size), $B_{\min}$ (contraction factor), and $B_{\max}$ (expansion factor) are chosen, and $\bar{h}$ is taken as

$$\bar{h} = \min\left\{ \max\{h_{\min}, B_{\min} h, \hat{h}\}, B_{\max} h, h_{\max} \right\}. \tag{47}$$

There are eight parameters in this process: $\bar{L}$, $\bar{R}$, $\bar{D}$, $h_{\min}$, $h_{\max}$, $B_{\min}$, $B_{\max}$, q. HOMPACK permits the user to specify nondefault values for any of these. The choice of $\bar{h}$ from (47) can be

refined further. If (39) converged in one iteration, then $\bar{h}$ should certainly not be smaller than h, hence set

$$\bar{h} := \max\{h, \bar{h}\} \tag{48}$$

if (39) only required one iteration.

To prevent divergence from the iteration (39), if (39) has not converged after K iterations, h is halved and a new prediction is computed. Every time h is halved the old value h_{old} is saved. Thus if (39) has failed to converge in K iterations sometime during this step, the new $\bar{h}$ should not be greater than the value h_{old} known to produce failure. Hence in this case

$$\bar{h} := \min\{h_{\text{old}}, \bar{h}\}. \tag{49}$$

Finally, if (39) required the maximum K iterations, the step size should not increase, so in this case set

$$\bar{h} := \min\{h, \bar{h}\}. \tag{50}$$

The logic in (48)–(50) is rarely invoked, but it does have a stabilizing effect on the algorithm.

The final phase, computation of the solution at $\lambda = 1$, begins when a point $P^{(2)}$ on γ is generated such that $P_1^{(2)} \geq 1$. The solution lies somewhere on γ between the previous point $P^{(1)}$ and $P^{(2)}$. The endgame now consists of iterating until convergence the sequence of steps: inverse interpolation with the Hermite cubic (38) for $\bar{s}$ such that $p(\bar{s})_1 = 1$; two iterations of (39) starting with $Z^{(0)} = p(\bar{s})$; replacing either $P^{(1)}$ or $P^{(2)}$ by $Z^{(2)}$ such that the solution on γ is always bracketed by $P^{(1)}$ and $P^{(2)}$. A precise statement of the endgame and the convergence criterion is given in [47].

3.3. Augmented Jacobian matrix algorithm. The augmented Jacobian matrix algorithm has four major phases: prediction, correction, step size estimation, and computation of the solution at $\lambda = 1$. The algorithm here is based on Rheinboldt [32], but with some significant differences: (1) a Hermite cubic rather than a linear predictor is used; (2) a tangent vector rather than a standard basis vector is used to augment the Jacobian matrix of the homotopy map; (3) updated QR factorizations and quasi-Newton updates are used rather than Newton's method; (4) different step size control, necessitated by the use of quasi-Newton iterations, is used; (5) a different scheme for locating the target point at $\lambda = 1$ is used which allows the Jacobian matrix of F to be singular at the solution $\bar{x}$ provided rank $D\rho_a(1, \bar{x}) = n$.

The prediction phase is exactly the same as in the normal flow algorithm. Having the points $P^{(1)} = (\lambda(s_1), x(s_1)), P^{(2)} = (\lambda(s_2), x(s_2))$ on γ with corresponding tangent vectors

$$T^{(1)} = \begin{pmatrix} \frac{d\lambda}{ds}(s_1) \\ \frac{dx}{ds}(s_1) \end{pmatrix}, \qquad T^{(2)} = \begin{pmatrix} \frac{d\lambda}{ds}(s_2) \\ \frac{dx}{ds}(s_2) \end{pmatrix}, \tag{51}$$

the prediction $Z^{(0)}$ of the next point on γ is given by (38).

In order to use this predictor, a means of calculating the tangent vector $T^{(2)}$ at a point $P^{(2)}$ is required. This is done by solving the system

$$\begin{bmatrix} D\rho_a\left(P^{(2)}\right) \\ T^{(1)\,t} \end{bmatrix} z = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{52}$$

for z, where $D\rho_a$ is the $n \times (n+1)$ Jacobian of ρ_a. Normalizing z gives

$$T^{(2)} = \frac{z}{\|z\|}. \tag{53}$$

The last row of (52) insures that the tangent $T^{(2)}$ makes an acute angle with the previous tangent $T^{(1)}$. It is the augmentation of the Jacobian matrix with this additional row which motivates the name "augmented Jacobian matrix algorithm." The solution to (52) is found by computing a QR factorization of the matrix, and then using back substitution [4].

Starting with the predicted point $Z^{(0)}$, the correction is performed by a quasi-Newton iteration defined by

$$Z^{(k+1)} = Z^{(k)} - \begin{bmatrix} A^{(k)} \\ T^{(2)\,t} \end{bmatrix}^{-1} \begin{pmatrix} \rho_a\left(Z^{(k)}\right) \\ 0 \end{pmatrix}, \qquad k = 0, 1, \ldots \tag{54}$$

where $A^{(k)}$ is an approximation to the Jacobian matrix $D\rho_a\left(Z^{(k)}\right)$. The last row of the matrix in (54) insures that the iterates lie in a hyperplane perpendicular to the tangent vector $T^{(2)}$. (54) is the quasi-Newton iteration for solving the augmented nonlinear system

$$\begin{pmatrix} \rho_a(y) \\ T^{(2)\,t}\left(y - Z^{(0)}\right) \end{pmatrix} = 0. \tag{55}$$

A corrector step $\Delta Z^{(k)}$ is the unique solution to the equation

$$\begin{bmatrix} A^{(k)} \\ T^{(2)\,t} \end{bmatrix} \Delta Z^{(k)} = \begin{pmatrix} -\rho_a\left(Z^{(k)}\right) \\ 0 \end{pmatrix}. \tag{56}$$

The matrix on the left side of this equation is produced by successive Broyden rank one updates [4] of the matrix in (52). Precisely, letting $Z^{(-1)} = P^{(2)}$, $A^{(-1)} = D\rho_a\left(P^{(2)}\right)$, and

$$M^{(k)} = \begin{bmatrix} A^{(k)} \\ T^{(2)\,t} \end{bmatrix},$$

the update formulas are

$$M^{(-1)} = \begin{bmatrix} A^{(-1)} \\ T^{(2)\,t} \end{bmatrix} = \begin{bmatrix} D\rho_a\left(P^{(2)}\right) \\ T^{(1)\,t} \end{bmatrix} + e_{n+1}\left(T^{(2)} - T^{(1)}\right)^t, \tag{57}$$

and

$$M^{(k+1)} = M^{(k)} + \frac{\left(\widetilde{\Delta}\rho_a - M^{(k)}\Delta Z^{(k)}\right)\Delta Z^{(k)\,t}}{\Delta Z^{(k)\,t}\,\Delta Z^{(k)}}, \qquad k = -1, 0, \ldots \tag{58}$$

where

$$\widetilde{\Delta}\rho_a = \begin{pmatrix} \rho_a\left(Z^{(k+1)}\right) - \rho_a\left(Z^{(k)}\right) \\ 0 \end{pmatrix}.$$

These updates can be done in QR factored form, requiring a total of $\mathcal{O}(n^2)$ operations for each iteration in the correction process [4]. When the iteration (54) converges within some tolerance, the final iterate $Z^{(*)}$ is accepted as the next point on the zero curve γ.

The step size estimation algorithm is an adaptation of a procedure developed by Rheinboldt [32]. At each point $P^{(k)}$ with tangent $T^{(k)}$ along γ, the curvature is estimated by the formula

$$\left\| w^{(k)} \right\| = \frac{2}{\Delta s_k} \left| \sin\left(\alpha_k/2\right) \right|, \tag{59}$$

where

$$w^{(k)} = \frac{T^{(k)} - T^{(k-1)}}{\Delta s_k}, \qquad \alpha_k = \arccos\left(T^{(k)\,t} T^{(k-1)}\right), \qquad \Delta s_k = \left\| P^{(k)} - P^{(k-1)} \right\|.$$

Intuitively, α_k represents the angle between the last two tangent vectors, and the curvature is approximated by the Euclidean norm of the difference between these two tangents divided by Δs_k.

This curvature data can be extrapolated to produce a prediction for the curvature for the next step

$$\hat{\xi}_k = \left\| w^{(k)} \right\| + \frac{\Delta s_k}{\Delta s_k + \Delta s_{k-1}} \left(\left\| w^{(k)} \right\| - \left\| w^{(k-1)} \right\| \right). \tag{60}$$

Since $\hat{\xi}_k$ can be negative, use

$$\xi_k = \max(\xi_{min}, \hat{\xi}_k) \quad \text{for some small} \quad \xi_{min} > 0, \tag{61}$$

as the predicted curvature for the next step.

The goal in estimating the optimal step size is to keep the error in the prediction $\|Z^{(0)} - Z^{(*)}\|$ relatively constant, so that the number of iterations required by the corrector will be stable. This is achieved by choosing the step size as

$$\hat{h} = \sqrt{\frac{2\delta_k}{\xi_k}}, \tag{62}$$

where δ_k represents the ideal starting error desired for the prediction step. δ_k is chosen as a function of the tolerance for tracking the curve and is also restricted to be no larger than half of Δs_k.

As with the normal flow algorithm, additional refinements on the optimal step size are made in order to prevent chattering and unreasonable values. In particular, $\bar{h}$ is chosen to satisfy equations (47) and (49). This $\bar{h}$ is then used as the step size for the next step.

The final phase of the algorithm, computation of the solution at $\lambda = 1$, is entered when a point $P^{(2)}$ is generated such that $P_1^{(2)} \geq 1$. $P^{(2)}$ is the first such point, so the solution must lie on γ somewhere between $P^{(2)}$ and the previous point $P^{(1)}$. The algorithm for finding this solution is a two step process which is repeated until the solution is found. First, starting from a point $P^{(k)}$, a prediction $Z^{(k-2)}$ for the solution is generated such that $Z_1^{(k-2)} = 1$. Second, a single quasi-Newton iteration is performed to produce a new point $P^{(k+1)}$ close to γ, but not necessarily on the hyperplane $\lambda = 1$.

Normally, the prediction $Z^{(k-2)}$ is computed by a secant method using the last two points $P^{(k)}$ and $P^{(k-1)}$:

$$Z^{(k-2)} = P^{(k)} + \left(P^{(k-1)} - P^{(k)}\right) \frac{\left(1 - P_1^{(k)}\right)}{\left(P_1^{(k-1)} - P_1^{(k)}\right)}. \tag{63}$$

However, this formula can potentially produce a disastrous prediction (e.g., if $|P_1^{(k-1)} - P_1^{(k)}| \ll |1 - P_1^{(k)}|$), so an additional scheme is added to ensure that this does not happen. In order to implement this scheme, a point $P^{(opp)}$ must be saved. This point is chosen as the last point computed from a quasi-Newton step which is on the opposite side of the hyperplane $\lambda = 1$ from $P^{(k)}$. Thus, the points $P^{(opp)}$ and $P^{(k)}$ bracket the solution. The prediction $Z^{(k-2)}$ may be bad whenever the inequality

$$\left\| Z^{(k-2)} - P^{(k)} \right\| > \left\| P^{(k)} - P^{(opp)} \right\| \tag{64}$$

is true. In this case, $Z^{(k-2)}$ is recomputed from the equation

$$Z^{(k-2)} = P^{(k)} + \left(P^{(opp)} - P^{(k)} \right) \frac{\left(1 - P_1^{(k)} \right)}{\left(P_1^{(opp)} - P_1^{(k)} \right)}. \tag{65}$$

This chord method, while much safer than the secant method (63), is used only in the special case (64) because it has a much slower rate of convergence than the secant method.

An exception to these linear prediction schemes occurs with the first step of the final phase. Since the tangents $T^{(1)}$ and $T^{(2)}$ at $P^{(1)}$ and $P^{(2)}$ are available, this information is used to generate a Hermite cubic polynomial $p(s)$ for calculating the first prediction point $Z^{(0)}$. This is done by finding the root $\bar{s}$ of the equation $p_1(s) = 1$. $Z^{(0)}$ is then given by

$$Z^{(0)} = p(\bar{s}). \tag{66}$$

After the predictor $Z^{(k-2)}$ has been determined, a quasi-Newton step is taken to get the point $P^{(k+1)}$. This step is defined by

$$P^{(k+1)} = Z^{(k-2)} + \Delta Z^{(k-2)}, \tag{67}$$

where $\Delta Z^{(k-2)}$ is the solution to (56). Again, the matrix in (56) is produced by the rank one updates (57) and (58).

The alternating process of computing a prediction and taking a quasi-Newton step is repeated until the solution is found.

3.4. HOMPACK organizational details. HOMPACK is organized in two different ways: by algorithm/problem type and by subroutine level. There are three levels of subroutines. The top level consists of drivers, one for each problem type and algorithm type. Normally these drivers are called by the user, and the user need know nothing beyond them. They allocate storage for the lower level routines, and all the arrays are variable dimension, so there is no limit on problem size. The second subroutine level implements the major components of the algorithms such as stepping along the homotopy zero curve, computing tangents, and the end game for the solution at $\lambda = 1$. A sophisticated user might call these routines directly to have complete control of the algorithm, or for some other task such as tracking an arbitrary parametrized curve over an arbitrary parameter range. The lowest subroutine level handles the numerical linear algebra, and includes some BLAS routines. All the linear algebra and associated data structure handling are concentrated in these routines, so a user could incorporate his own data structures by writing his own versions of these low level routines. Also, by concentrating the linear algebra in subroutines, HOMPACK can be easily adapted to a vector or parallel computer.

Table 1. Taxonomy of homotopy subroutines.

$x = f(x)$		$F(x) = 0$		$\rho(a, \lambda, x) = 0$		algorithm
dense	sparse	dense	sparse	dense	sparse	
FIXPDF	FIXPDS	FIXPDF	FIXPDS	FIXPDF	FIXPDS	ordinary differential equation
FIXPNF	FIXPNS	FIXPNF	FIXPNS	FIXPNF	FIXPNS	normal flow
FIXPQF	FIXPQS	FIXPQF	FIXPQS	FIXPQF	FIXPQS	augmented Jacobian matrix

The organization of HOMPACK by algorithm/problem type is shown in Table 1, which lists the driver name for each algorithm and problem type.
The naming convention is

$$FIXP \left\{ \begin{matrix} D \\ N \\ Q \end{matrix} \right\} \left\{ \begin{matrix} F \\ S \end{matrix} \right\},$$

where $D \approx$ ordinary differential equation algorithm, $N \approx$ normal flow algorithm, $Q \approx$ augmented Jacobian matrix algorithm, $F \approx$ dense Jacobian matrix, and $S \approx$ sparse Jacobian matrix. Using brackets to indicate the three subroutine levels described above, the natural grouping of the HOMPACK routines is:

[FIXPDF] [FODE, ROOT, SINTRP, STEPS] [DCPOSE]

[FIXPDS] [FODEDS, ROOT, SINTRP, STEPDS] [GMFADS, MFACDS, MULTDS, PCGDS, QIMUDS, SOLVDS]

[FIXPNF] [ROOTNF, STEPNF, [TANGNF]] [ROOT]

[FIXPNS] [ROOTNS, STEPNS, TANGNS] [GMFADS, MFACDS, MULTDS, PCGDS, PCGNS, QIMUDS, ROOT, SOLVDS]

[FIXPQF] [ROOTQF, STEPQF, TANGQF] [QRFAQF, QRSLQF, R1UPQF, UPQRQF]

[FIXPQS] [ROOTQS, STEPQS, TANGQS] [GMFADS, MULTDS, PCGQS, SOLVDS]

The BLAS subroutines used by HOMPACK are DAXPY, DCOPY, DDOT, DNRM2, DSCAL, D1MACH, IDAMAX.

The user written subroutines, of which exactly two must be supplied depending on the driver chosen, are F, FJAC, FJACS, RHO, RHOA, RHOJAC, RHOJS.

The special purpose polynomial system solver POLSYS is essentially a high level driver for HOMPACK. POLSYS requires special versions of RHO and RHOJAC (subroutines normally provided by the user). These special versions are included in HOMPACK, so for a polynomial system the user need only call POLSYS, and define the problem directly to POLSYS by specifying the polynomial coefficients. POLSYS scales and computes partial derivatives on its own. Thus the user interface to POLSYS and HOMPACK is clean and simple. The only caveat is that FFUNP cannot recognize patterns of repeated expressions in the polynomial system, and so may be less efficient than a hand crafted version. If great efficiency is required, the user can modify the default FFUNP; the sections in the code which must be changed are clearly marked. The grouping is:

[POLSYS] [POLYNF, POLYP, ROOTNF, STEPNF, TANGNF] [DIVP, FFUNP, GFUNP, HFUNP, HFUN1P, INITP, MULP, OTPUTP, POWP, RHO, RHOJAC, ROOT, SCLGNP, STRPTP]

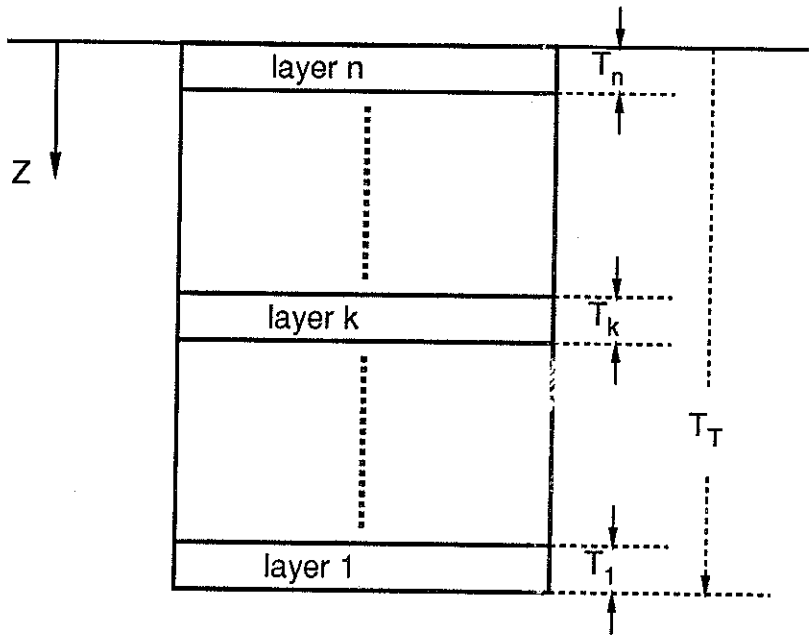


Figure 2. Geometry of half of a $2n$-layered symmetric laminate.

4. Engineering applications.

4.1. Optimal composite plate design. Composite materials are ideal for structural applications where high strength-to-weight and stiffness-to-weight ratios are required. Design optimization of composite structures has gained importance in recent years as the engineering applications of fiber-reinforced materials have increased and weight savings has become an essential design objective, especially for aircraft and spacecraft structures. The laminates considered here are symmetric about the middle surface with $2n$ layers (see Figure 2), so that the bending response is not coupled to the membrane action. The optimization problem is to maximize the buckling load of a $2n$-layered composite plate (Figure 3) for a given total plate thickness. The thickness of each layer is assumed to be constant over the plate, and for a given stacking sequence of the ply orientations, each thickness is taken as a design variable.

This is an instance of a general engineering design problem, namely to maximize the lowest buckling load of a structure for a given amount of resources. The structure is discretized by finite elements. Expressing the lowest buckling load with Rayleigh's quotient, the problem is written as

$$\begin{aligned} &\max_{v} \min_{u} \frac{u^T K u}{u^T K_G u} \\ &\text{such that} \quad c^T v - \theta = 0 \\ &\qquad \text{and} \quad v_{i\,min} \le v_i \le v_{i\,max} \quad \text{for } i = 1, \ldots, M, \end{aligned} \tag{68}$$

where v is a vector of design variables with components v_i, u is the displacement vector, K and K_G (depending on v) are the stiffness matrix and the geometric stiffness matrix, respectively, c is a positive cost vector, and θ is the amount of available resources. The M design variables are subject to upper and lower bounds, $v_{i\,max}$ and $v_{i\,min}$, respectively.

A typical optimization method, applied to solve this problem, starts from a given design and continuously searches for better designs until it finds an optimum design. The trial designs along

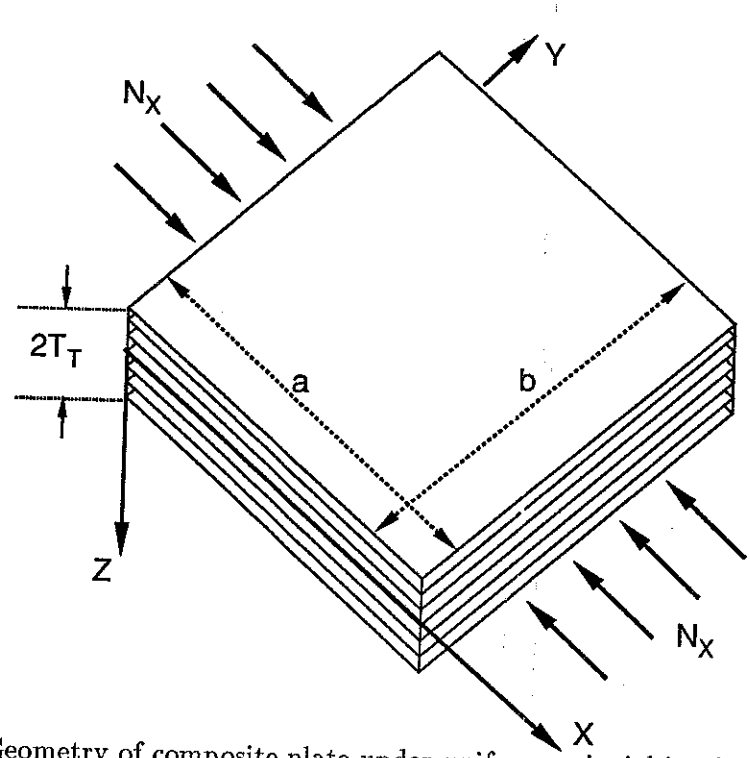


Figure 3. Geometry of composite plate under uniform uniaxial in-plane load.

the path are of no value. The proposed method instead proceeds along a path of optimal designs for increasing amounts of resource θ. The resource θ is varied between the minimum θ_{min} required to satisfy the lower bound constraints and a maximum θ_{max} when all variables are at their upper bounds.

The path consists of several smooth segments, each segment being characterized by a set I_A of variables which are at their upper or lower bounds. Along each segment, some inequality constraints can be treated as equality constraints,

$$v_j = v_{j\,min} \quad \text{or} \quad v_j = v_{j\,max} \quad \text{for } j \in I_A, \tag{69}$$

so that these variables can be eliminated from the optimization problem, while the other variables do not have to be constrained. The optimization problem along a segment can, therefore, be written as

$$\begin{aligned} &\max_{v_i} \min_{u} \frac{u^T K u}{u^T K_G u} \quad \text{for } i \notin I_A \\ &\text{such that} \quad c^T v - \theta = 0. \end{aligned} \tag{70}$$

The solution of the problem consists of three related problems: solving the optimization problem along a segment, locating the end of the segment where the set I_A changes, and finding the set I_A for the next segment.

It is common practice to normalize the displacement vector u such that the denominator of Rayleigh's quotient is unity and to treat this as an equality constraint. Then, using Lagrange multipliers η and μ, the augmented function P^* is formed:

$$P^* = u^T K u - \eta \left[u^T K_G u - 1\right] - \mu \left[c^T v - \theta\right]. \tag{71}$$

The following stationary conditions are obtained by taking the first derivative of P^* with respect to v_i, u, η, and μ, and setting it equal to zero:

i) Optimality conditions

$$u^T \frac{\partial K}{\partial v_i} u - \eta u^T \frac{\partial K_G}{\partial v_i} u - \mu c_i = 0 \quad \text{for } i \notin I_A. \tag{72}$$

ii) Stability conditions

$$Ku - \eta K_G u = 0. \tag{73}$$

iii) Normalization constraint

$$1 - u^T K_G u = 0. \tag{74}$$

iv) Total resource constraint

$$\theta - c^T v = 0. \tag{75}$$

Equations (72)–(75) form a system of nonlinear equations to be solved for v_i, u, η, and μ. A homotopy method is used to find the solutions of these equations as a function of θ.

In certain ranges of structural resources, the optimal solution is known to be bimodal, i.e., the lowest buckling load is a repeated eigenvalue. The formulation for bimodal solutions is given in the appendix of [35]. The existence of bimodal solutions also introduces additional transitions (bimodal to unimodal and vice versa) along the path of optimum solutions.

The homotopy method as described here earlier is intended to solve a *single* nonlinear system of equations, and converge from an arbitrary starting point with probability one. In this context $\theta \in [0, 1]$, and the zero curve γ is bounded and leads to the (single) desired solution at $\theta = 1$. The a vector, viewed as an artificial perturbation of the problem, plays a crucial role. In the version of the method employed here, $\theta \in (\theta_0, \theta_1)$, *each* point along γ has physical significance, and a is fixed at zero (no perturbation). Because a is not random, the claimed properties for γ hold only in subintervals (θ_0, θ_1) of $[0, \infty)$. Detecting and dealing with these subinterval transition points is the essence of the modification of the homotopy method used in this section.

Switching from one segment to the next

There are four types of events which end a segment and start a new one:
Type 1: a bound constraint becoming active (i.e., being satisfied as an equality);
Type 2: a bound constraint becoming inactive;
Type 3: transition from a unimodal solution to a bimodal solution;
Type 4: transition from a bimodal solution to a unimodal solution.

To switch from one segment to the next, we first need to locate the transition point. At a transition point there are a number of solution paths which satisfy the stationary equations, and we need to choose the optimum path.

Transition points are located by checking the bound constraints and the optimality conditions. The bound constraints

$$v_{i\,min} \le v_i \le v_{i\,max} \quad \text{for } i = 1, \ldots, M \tag{76}$$

are checked to detect a transition point of type 1.

Optimality of the solution is checked by the Kuhn-Tucker conditions and the second-order conditions discussed below. The solution satisfies the Kuhn-Tucker conditions when all Lagrange multipliers are nonnegative. So a transition of type 2 is detected by checking the positivity of the

Lagrange multipliers associated with the bound constraints. These multipliers are obtained by adding the bound constraints to the formulation (70) and replacing the augmented function P^* by

$$P^* = u^T K u - \eta \left[u^T K_G u - 1\right] - \mu \left[c^T v - \theta\right] - \sum_{i \in I_A} \lambda_{1i} \left[v_{i\,min} - v_i\right] - \sum_{i \in I_A} \lambda_{2i} \left[v_i - v_{i\,max}\right]. \quad (77)$$

Taking the first derivative of P^* with respect to v_i gives

$$u^T \frac{\partial K}{\partial v_i} u - \eta u^T \frac{\partial K_G}{\partial v_i} u - \mu c_i + \lambda_{1i} - \lambda_{2i} = 0 \quad \text{for } i \in I_A. \quad (78)$$

Since λ_{1i} is 0 for $v_i \neq v_{i\,min}$ and λ_{2i} is 0 for $v_i \neq v_{i\,max}$ for the above equations, λ_{1i} and λ_{2i} are given by

$$\begin{aligned} \lambda_{1i} &= -u^T \frac{\partial K}{\partial v_i} u + \eta u^T \frac{\partial K_G}{\partial v_i} u + \mu c_i \quad \text{for } v_i = v_{i\,min} \\ \lambda_{2i} &= u^T \frac{\partial K}{\partial v_i} u - \eta u^T \frac{\partial K_G}{\partial v_i} u - \mu c_i \quad \text{for } v_i = v_{i\,max}. \end{aligned} \quad (79)$$

A type 2 transition is detected by a Lagrange multiplier becoming nonpositive. Similar equations for the bimodal case are given in the appendix of [35].

The bimodal formulation replaces η by η_1 and η_2 which are the Lagrange multipliers for the normalization constraints on the two buckling modes. When one of them becomes negative, the corresponding mode should be removed for the optimum design, so that we have a transition of type 4 from bimodal to unimodal design.

For a transition of type 3, we need to check if there is another buckling mode associated with a lower buckling load. This can be accomplished by checking the second-order optimality conditions for the buckling mode variables u given by

$$r^T \left[\nabla_u^2 P^*\right] r > 0 \quad \text{for every } r \text{ such that } \quad \nabla_u h^T r = 0 \quad (80)$$

where

$$\begin{aligned} \left[\nabla_u^2 P^*\right] &= \left[\frac{\partial^2 P^*}{\partial u_s \partial u_t}\right] \\ \nabla_u h &= \left[\frac{\partial h}{\partial u_s}\right] \\ h &= u^T K_G u - 1. \end{aligned}$$

Alternatively we can solve the buckling problem (73) for the current design and check whether the buckling load obtained from the stationary conditions is truly the lowest one. The transition of type 3 is detected by checking if

$$p \neq p_1 \quad (81)$$

where p is the buckling load obtained from the stationary conditions while p_1 is the first buckling load obtained by solving the stability conditions (73) for the given structure.

Once a transition point is located, we need to choose a path which satisfies the optimality conditions. Choosing an optimum path constitutes finding a set of active bound constraints for type 1 and 2 transitions and the correct buckling modes for type 3 and 4 transitions. These are obtained by using the Lagrange multipliers of the previous path and the sensitivity calculation on the buckling load. The procedure is explained separately for each type of transition.

A type 1 transition occurs when one of design variables, v_i, hits the upper or lower bound. Then v_i is set at $v_{i\,max}$ or $v_{i\,min}$ and treated as a constant value. The number of design variables is reduced by one.

At a type 2 transition, one of the Lagrange multipliers for the bound constraints, λ_{1i} and λ_{2i}, is found to be negative. The bound constraint corresponding to the negative λ_{1i} or λ_{2i} is set to be inactive and the number of design variables is increased by one.

At a transition from a unimodal solution to a bimodal solution (a type 3 transition), the formulation requires two buckling modes, u_1 and u_2, for the solution of the upcoming bimodal path. These modes can be obtained by solving the stability conditions (73) of the previous unimodal formulation, since the stability conditions give two buckling modes at the bimodal transition point.

At a transition from a bimodal to a unimodal solution (a type 4 transition), two buckling modes are given from the bimodal solution. One of the Lagrange multipliers for the normalization constraints, η, is known to be negative from the previous transition check, so the buckling mode corresponding to the positive η is chosen.

Some of the above transitions can occur simultaneously. Special treatment is required in certain cases where the Lagrange multipliers are not available. In general, the optimum design requires at least one design variable v_i for a unimodal case and two design variables for a bimodal case. At a type 1 transition, the number of design variables is reduced by one, and at a type 3 transition the bimodal formulation requires one more design variable in case the previous unimodal path has only one design variable. So some type 1 or type 3 transitions occur simultaneously with a type 2 transition which allows an additional design variable. In that case, the Lagrange multipliers λ_{1i} and λ_{2i}, which are used at a type 2 transition to determine a new design variable, are not available. We then rely on the sensitivity information of p with respect to v. For a unimodal case, the location of the new design variable v_i is determined where $dp/d\theta$ is maximized. For a bimodal case, we need to find a combination of i and j which maximizes the value of the bimodal buckling load for a small increment of the total available resource. Considering the bound constraints in the formulation, the new design variables are determined by

$$\max_{i,j} \frac{dp}{d\theta} = \frac{\partial p_1}{\partial v_i}\frac{dv_i}{d\theta} + \frac{\partial p_1}{\partial v_j}\frac{dv_j}{d\theta} \tag{82}$$

such that

$$\frac{\partial p_1}{\partial v_i}\frac{dv_i}{d\theta} + \frac{\partial p_1}{\partial v_j}\frac{dv_j}{d\theta} = \frac{\partial p_2}{\partial v_i}\frac{dv_i}{d\theta} + \frac{\partial p_2}{\partial v_j}\frac{dv_j}{d\theta}$$

$$\frac{dv_i}{d\theta} \geq 0 \quad \text{for } v_i = v_{i\,min}$$

$$\frac{dv_i}{d\theta} \leq 0 \quad \text{for } v_i = v_{i\,max}$$

$$\frac{dv_j}{d\theta} \geq 0 \quad \text{for } v_j = v_{j\,min}$$

$$\text{and } \frac{dv_j}{d\theta} \leq 0 \quad \text{for } v_j = v_{j\,max}$$

where p_1 and p_2 are the buckling loads corresponding to the buckling modes u_1 and u_2, respectively.

After we obtain the design variables v and the buckling modes u, we need the Lagrange multipliers μ and η at the transition point to complete the set of starting values for the next solution path. These are obtained by solving the stationary conditions for the given u and v. For

example, in the unimodal case, η is obtained from the stability conditions (73) and μ is obtained by solving one of the optimality conditions (72).

Summary

A typical optimization method starts from a given design and continuously searches for better designs until it finds an optimum design. The trial designs along the path are of no value. Here a strategy for tracing a path of optimum solutions parameterized by an amount of available resources was discussed. Equations for the optimum path were obtained using Lagrange multipliers, and were solved by a homotopy method.

The solution path has several branches due to changes in the active constraint set and transition from unimodal to bimodal solutions. The Lagrange multipliers and the second-order optimality conditions were used to detect branching points and to switch to the optimum solution path.

In [35] this procedure was applied to the design of a foundation which supports a column for maximum buckling load, where the total available foundation was used as a homotopy parameter. Starting from a minimum foundation which satisfies the lower bound (in this case zero), a set of optimum foundation designs was obtained for the full range of total foundation stiffness. Numerical results for the design of composite plates described here, where the total plate thickness is the resource parameter being varied, are in [36].

4.2. Fuel-optimal orbital rendezvous problem. The problem is to find a minimum fuel rendezvous trajectory between two bodies, the non-maneuvering target and the interceptor. The interceptor trajectory consists of Keplerian coasting arcs separated by impulsive thrusting, characterized by a change in velocity (magnitude and direction) A final impulse is applied at the end of the interceptor trajectory to provide a velocity match with the target. Hence the number of impulses equals the number of coasting arcs. The maneuver must be completed within some specified time and the trajectory must avoid passing too near the earth, i.e., the arcs must not violate a minimum radius constraint. The fuel-optimal problem translates to minimizing the total change in the velocity (characteristic velocity).

The notation used is:

η	– change in true anomaly,
$\vec{r}(\eta)$	– radius vector,
$\hat{r}(\eta)$	– unit vector in the radial direction,
u	– reciprocal of the magnitude of the radius vector,
$\vec{v}(\eta)$	– velocity vector,
$h(\eta)$	– magnitude of the angular momentum vector,
$\hat{h}(\eta)$	– unit vector in the direction of angular momentum.

The variables are the coasting angles on each arc including a possible initial coast, the components of the velocity change vector, and the coasting angle of the target. The forward equations of motion for any subarc are:

$$u(\eta) = \frac{\mu}{h^2} + \left(u(0) - \frac{\mu}{h^2}\right)\cos\eta + u'(0)\sin\eta,$$
$$\hat{r}(\eta) = \hat{r}(0)\cos\eta + \hat{r}'(0)\sin\eta,$$

with time of flight

$$T(\eta) = \int_0^{\eta} \frac{1}{hu^2(\theta)} d\theta.$$

The constraints are:

final position match	$\vec{r}_f - \vec{r}_t(\eta_t) = 0,$
final velocity match	$\vec{v}_f - \vec{v}_t(\eta_t) = 0,$
time of flight constraint	$T_f - T_t = 0,$
nonnegativity of the coasting arcs of the interceptor	$\eta_i \geq 0, \quad i = 1, \ldots, nim,$
nonnegativity of the coasting arc of the target	$\eta_t \geq 0,$
time limit for rendezvous	$T_{\max} - T_f \geq 0,$
minimum radius constraint for each coasting arc ...	
except the initial coast arc of the interceptor ..	$u_0 - u_{j\max} \geq 0, \quad j = 1, \ldots, nim - 1,$
nonnegativity of the radius constraint	$u_{j\min} \geq 0, \quad j = 1, \ldots, nim - 1.$

The subscript f refers to the conditions on the interceptor trajectory after the final impulse and the subscript t refers to conditions on the target. nim is the number of impulses. The value of $u_{j\max}$ in these constraints is given by the rather awkward and difficult to compute expression

$$\frac{1}{u_{\max}} = \begin{cases} \text{perigee radius, if perigee passage occurs on subarc,} \\ \min(r_{initial}, r_{final}), \text{ otherwise.} \end{cases}$$

The optimization problem, subject to all the above constraints, is

$$\min_S V(x),$$

where

$$S = \{\eta_t, (\eta, \Delta u', \Delta h, \phi)_j,\ j = 1, \ldots, nim\},$$

and

$$V = \sum_{j=1}^{nim} \sqrt{u_{j+1}^2(0)\left[h_{j+1}^2 - 2h_j h_{j+1}\cos\phi_j + h_j^2\right] + \left[\Delta h_j u'_{j+1}(0) + \Delta u'_j h_j\right]^2}.$$

For u, u', and h, the subscript j denotes the conditions at the beginning of the jth subarc, and on the variables $\Delta u'$, Δh, and ϕ the subscript j denotes the jth impulse which occurs at the end of the jth subarc.

Using the formulation of equations (29) and (32), numerous such rendezvous problems have been solved, both in-plane and out-of-plane, and with 2, 3, 4, or 5 impulses. See [37] for more details.

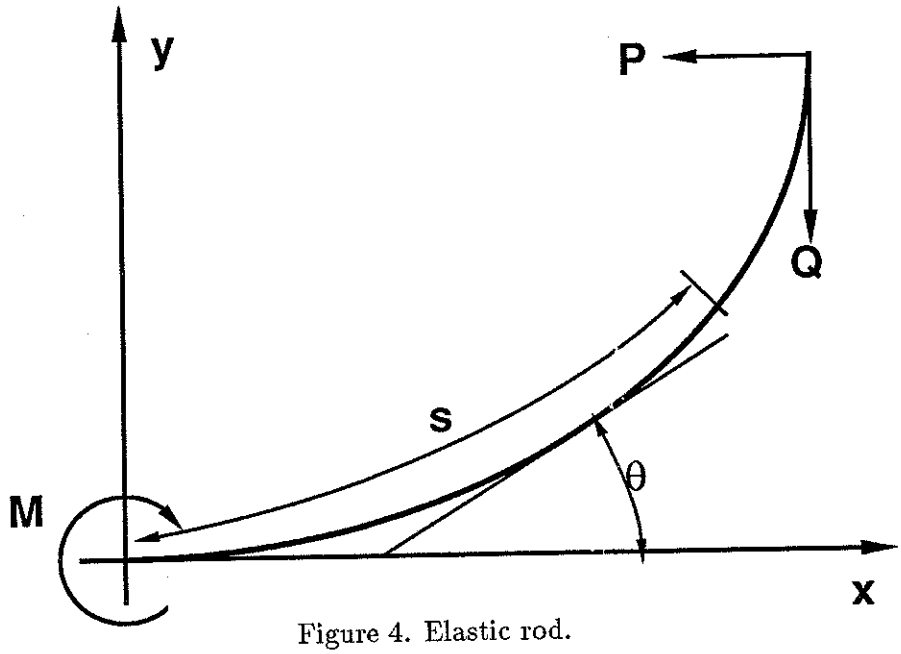


Figure 4. Elastic rod.

4.3. Elastic rod. Consider a thin incompressible elastic rod clamped at the origin and acted on by forces Q, P and torque M (see Figure 4). The governing nondimensionalized equations are

$$\frac{dx}{ds} = \cos\theta, \quad \frac{dy}{ds} = \sin\theta, \quad \frac{d\theta}{ds} = Qx - Py + M, \tag{83}$$

$$x(0) = y(0) = \theta(0) = 0, \tag{84}$$

$$x(1) = a, \quad y(1) = b, \quad \theta(1) = c. \tag{85}$$

The cantilever beam problem, which is to find the position (a, b) of the tip of the rod given the forces $Q \neq 0$ and $P = 0$, has a closed-form solution in terms of elliptic integrals. The inverse problem, where the a, b, c are specified and Q, P, M are to be determined, has no similar closed-form solution. This inverse problem is ferociously nonlinear and extremely difficult. For large deformations, $c = 6\pi$ for example, the rod is wound like a coil spring and its shape is very sensitive to small perturbations in Q, P, or M.

Let

$$v = \begin{pmatrix} Q \\ P \\ M \end{pmatrix} \tag{86}$$

and $x(s; v)$, $y(s; v)$, $\theta(s; v)$ be the solution (dependent on v) of the initial value problem (83)–(84). Then an equivalent formulation of the nonlinear two-point boundary value problem (83)–(85) is to find a vector v such that

$$F(v) = \begin{pmatrix} x(1; v) - a \\ y(1; v) - b \\ \theta(1; v) - c \end{pmatrix} = 0. \tag{87}$$

This particular formulation is based on shooting. Nonlinear systems different from (87) could be derived based on multiple shooting, finite differences, polynomial or spline collocation, spectral methods, or Galerkin methods. The best way to approximate the solution to (83)–(85) is not the issue here. The issue is how to solve the particular given nonlinear system of equations (87).

Newton and quasi-Newton methods, even the very best such as HYBRJ from Argonne's MINPACK package [23], fail dismally when applied to (87). A simple continuation scheme, such as tracking the zeros of

$$\rho_w(\lambda, v) = \lambda F(v) + (1-\lambda)(v-w) \tag{88}$$

as λ is increased from 0 to 1, also fails. The zero curve of $\rho_w(\lambda, v)$ in (88) emanating from $(0, w)$ diverges to infinity. In fact this divergence also occurs if $F(v)$ in (88) is replaced by $DF(v)$ for any diagonal orthogonal matrix D. Nonlinear least squares algorithms attempting to minimize $F(v)^t F(v)$ also quickly fail, since there are too many local minima $\tilde{v}$ that are not global minima ($F(\tilde{v}) \neq 0$).

Consider the function (known as a *homotopy map*) $\rho : E^3 \times [0,1) \times E^3 \to E^3$ defined by

$$\rho(d, \lambda, v) = \rho_d(\lambda, v) = \begin{pmatrix} x(1;v) - [\lambda a + (1-\lambda)d_1] \\ y(1;v) - [\lambda b + (1-\lambda)d_2] \\ \theta(1;v) - [\lambda c + (1-\lambda)d_3] \end{pmatrix}. \tag{89}$$

ρ_d is a *homotopy* (the technical term from topology) because it continuously deforms one function (in this case $\rho_d(0, v)$) to another function (in this case $\rho_d(1, v) = F(v)$). Note that $\rho(d, \lambda, v)$ is C^2 and that its Jacobian matrix

$$D\rho(d, \lambda, v) = \left[-(1-\lambda)I,\ -(a,b,c)^t + d,\ D_v F(v)\right] \tag{90}$$

has full rank (rank $= 3$) on $\rho^{-1}(0)$. In differential geometry jargon, ρ is said to be *transversal to zero*. The mathematics then says that for almost all vectors $d \in E^3$ (in the sense of Lebesgue measure), the map ρ_d is also transversal to zero. What this means geometrically is that the zero set of ρ_d consists of smooth disjoint curves that do not intersect themselves or bifurcate, and have endpoints only at $\lambda = 0$ or $\lambda = 1$ (see Figure 1). In general under suitable conditions (described in Section 2) there is a zero curve γ of $\rho_d(\lambda, v)$ stretching from a known solution v_0 at $\lambda = 0$ to the desired solution $\bar{v}$ at $\lambda = 1$.

For the homotopy map (89) such a zero curve γ does indeed exist, and a solution $\bar{v}$ to (87) can be found by tracking γ starting from $(0, v_0)$. v_0 and d are related by $x(1; v_0) = d_1$, $y(1; v_0) = d_2$, $\theta(1; v_0) = d_3$. Since the theory says that everything works for almost all d, by the implicit function theorem, everything also works for almost all v_0. Thus in practice v_0 was chosen at random and d computed from v_0, rather than vice versa. More details on this example can be found in [57].

4.4. Heavy elastic sheets. The overhang of a semi-infinite elastic sheet over a corner is important in structural engineering and the textile and paper industries. Figure 5 shows such an elastic sheet freely resting on a semi-infinite rigid foundation at $x' \geq 0$. Due to the weight of the overhang, the sheet is raised and separated from the foundation in the segment from the corner 0 to the point of contact at $x' = x'_c$. We assume the corner offers little frictional resistance. The sheet is kept

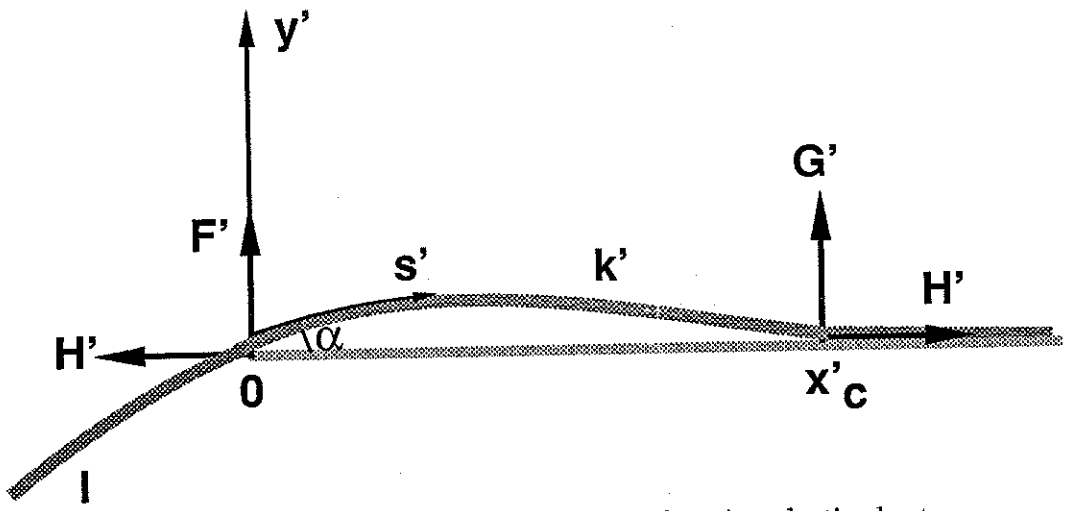


Figure 5. Coordinate system and forces for overhanging elastic sheet.

in equilibrium by the horizontal force H' at x'_c. This horizontal force may be due to frictional resistance of the semi-infinite segment of contact $x' \geq x'_c$.

Let s' be the arc length from 0 and l be the length of the overhang. The sheet can be divided into three segments: the overhang from $s' = -l$ to $s' = 0$, the raised segment from $s' = 0$ to $s' = k'$, and a contact segment $s' \geq k'$ (where $x' \geq x'_c$). Since the force must be normal to the sheet at the point 0, the vertical force there (F') is related to H' by

$$\tan\alpha = \frac{H'}{F'},$$

where α is the angle of inclination at 0. If ρ is the weight per unit length, the vertical force G' at the point of contact $s' = k'$ is then

$$G' = (l + k')\rho - F'.$$

A local balance of momentum (Figure 5) gives, for the overhang segment,

$$m + dm = m - \rho(l + s')\cos\theta\, ds'. \tag{91}$$

Here m is the local moment, and θ is the local angle of inclination. If the sheet is thin enough, the local moment is proportional to the local curvature:

$$m = EI\frac{d\theta}{ds'}, \tag{92}$$

where EI is the flexural rigidity. We normalize all lengths by l and drop primes. Equations (91), (92) become

$$\frac{d^2\theta}{ds^2} = -K(1 + s)\cos\theta, \tag{93}$$

where $K = \rho l^3/EI$ represents the relative importance of density and length to flexural rigidity. The boundary conditions are

$$\theta(0) = \alpha, \qquad \frac{d\theta}{ds}(0) = \lambda, \tag{94}$$

$$\frac{d\theta}{ds}(-1) = 0. \tag{95}$$

Similarly, the equation for the raised segment is

$$\frac{d^2\theta}{ds^2} = \big[F - K(1+s)\big]\cos\theta + F\tan\alpha\sin\theta. \tag{96}$$

Here all forces have been normalized by EI/l^2. The shape of the sheet is given by

$$\frac{dx}{ds} = \cos\theta, \qquad \frac{dy}{ds} = \sin\theta \tag{97}$$

with the boundary conditions

$$x(0) = y(0) = 0, \qquad \theta(0) = \alpha, \qquad \frac{d\theta}{ds}(0) = \lambda, \tag{98}$$

$$y(k) = \theta(k) = \frac{d\theta}{ds}(k) = 0. \tag{99}$$

Given K, Equations (93)–(99) are to be solved concurrently for the unknowns α, λ, F, and k.

An asymptotic solution is possible for small K. For general K the deflections are no longer small and numerical integration is necessary. Define

$$v = (\alpha, \lambda, F, k) \tag{100}$$

and let $x(s;v)$, $y(s;v)$, $\theta(s;v)$ be the solution to the initial value problem Equations (93), (96), (97) with the initial conditions (94), (98), (100). Then the original two-point boundary value problem is equivalent to

$$f(v) = \left(y(k;v),\ \theta(k;v),\ \frac{d\theta}{ds}(k;v),\ \frac{d\theta}{ds}(-1;v)\right) = 0. \tag{101}$$

Equation (101) can be solved by a homotopy method similar to that described in Section 2.3.1. The algorithm requires the Jacobian matrix $Df(v)$ of $f(v)$, and the partial derivatives $\dfrac{\partial f}{\partial v_i}(v)$. These are computed as follows:

Set $z_1 = x$, $z_2 = y$, $z_3 = \theta$, $z_4 = \theta' = d\theta/ds$, $z_5 = \partial x/\partial v_i$, $z_6 = \partial y/\partial v_i$, $z_7 = \partial\theta/\partial v_i$, $z_8 = \partial\theta'/\partial v_i$ and consider the differential equations

$$\begin{aligned}
z_1' &= \cos z_3,\\
z_2' &= \sin z_3,\\
z_3' &= z_4,\\
z_4' &= -K(1+s)\cos z_3 + F(\cos z_3 + \tan\alpha\sin z_3),\\
z_5' &= -z_7\sin z_3,\\
z_6' &= z_7\cos z_3,\\
z_7' &= z_8,\\
z_8' &= K(1+s)z_7\sin z_3 + T,
\end{aligned} \tag{102}$$

where

$$T = \frac{\partial}{\partial v_i}\big(F(\cos z_3 + \tan\alpha\sin z_3)\big) \tag{103}$$

has a different form depending on v_i. For $v_1 = \alpha$, the initial conditions are

$$z(0) = (0, 0, \alpha, \lambda, 0, 0, 1, 0); \tag{104}$$

for $v_2 = \lambda$

$$z(0) = (0, 0, \alpha, \lambda, 0, 0, 0, 1); \tag{105}$$

for $v_3 = F$

$$z(0) = (0, 0, \alpha, \lambda, 0, 0, 0, 0); \tag{106}$$

for $v_4 = k$

$$z(0) = (0, 0, \alpha, \lambda, 0, 0, 0, 0). \tag{107}$$

Thus solving the initial value problem given by Equations (102) and (104) produces, e.g., $\frac{\partial y}{\partial \alpha}(k)$, which is the $(1,1)$ entry in the Jacobian matrix $Df(v)$. Using the differential equation (102) with $T = 0$ and initial conditions (104) or (105) produces the partials of $\theta'(-1)$, where the initial value problem is solved backwards from $s = 0$ to $s = -1$. Since the differential equation for $s \leq 0$ does not depend on F or k,

$$\frac{\partial \theta'}{\partial F}(-1) = \frac{\partial \theta'}{\partial k}(-1) = 0.$$

These initial value problems were solved by a variable step, variable order ODE code which is accurate, efficient, and robust [34], [43], [61], [62]. The combination of a globally convergent homotopy method and a sophisticated ODE method proves to be very successful on this problem.

4.5. Heavy elastic cylindrical shells. Important construction problems in outer space and undersea involve heavy elastic cylinders. Depending on the rigidity of the elastic wall material, the cylinder may collapse under its own weight. There are four distinct cases, governed by a nondimensional parameter B (see Figure 6). Starting from a perfect cylinder ($B = 0$), as B increases the point contact (case 1) widens to a line contact (case 2); then the top sags until it touches the bottom for a point-line contact (case 3); then ultimately the top also makes a line contact with the bottom (case 4). The governing equations for all four cases are

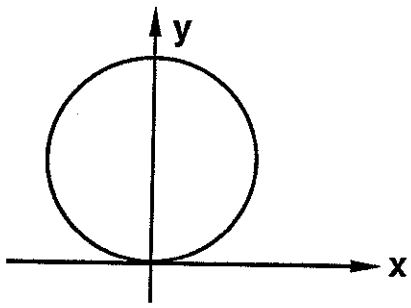
$$\frac{dx}{ds} = \cos\theta, \qquad \frac{dy}{ds} = \sin\theta,$$
$$\frac{d^2\theta}{ds^2} = A\sin\theta + (C - Bs)\cos\theta.$$

For case 1, $C = B$ and the boundary conditions are

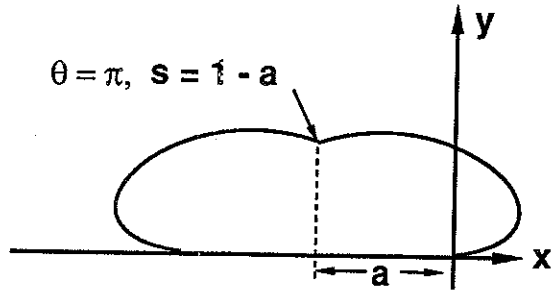
$$x(0) = y(0) = \theta(0) = 0,$$
$$x(1) = 0, \qquad \theta(1) = \pi.$$

For case 2, $C = B(1 - a)$ and the boundary conditions are

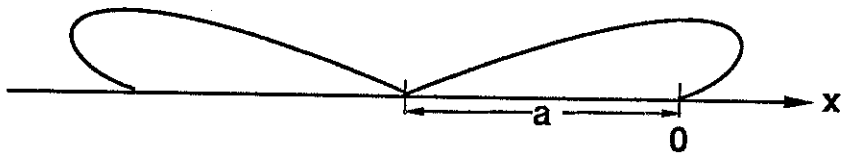
$$x(0) = y(0) = \theta(0) = \theta'(0) = 0,$$
$$x(1 - a) = -a, \qquad \theta(1 - a) = \pi.$$



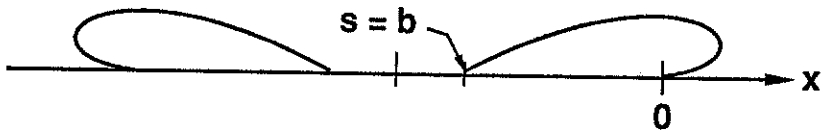
Case 1: point contact



Case 2: line contact



Case 3: line-point contact



Case 4: line-line contact

Figure 6. Heavy elastic cylindrical shells.

For case 3, the boundary conditions are

$$x(0) = y(0) = \theta(0) = \theta'(0) = 0,$$
$$x(1-a) = -a, \qquad y(1-a) = 0, \qquad \theta(1-a) = \pi.$$

For case 4, the boundary conditions are

$$x(0) = y(0) = \theta(0) = \theta'(0) = 0,$$
$$y(b) = 0, \qquad \theta'(b) = 0.$$

For cases 1 and 2, quasi-Newton methods are adequate and efficient if a good computer code is used. For cases 3 and 4, where B is large, quasi-Newton methods are feasible but very expensive because of their small domain of practical application. If the starting point is too far away from the solution, quasi-Newton codes such as HYBRJ from Argonne's MINPACK software package fail to make progress toward the solution and give an error return. The homotopy map

$$\rho_a(\lambda, v) = \lambda F(v) + (1-\lambda)(v-a),$$

where v consists of the appropriate initial conditions and parameters (depending on the case) and $F(v)$ is defined by shooting, works very well for large B [43]. This is an example of a problem on which quasi-Newton methods do not totally fail, and yet the homotopy algorithm is more efficient.

4.6. Porous channel flow in a rotating system. Lubrication in rotating machinery and flow under the polar ice cap are examples of porous-channel flow in a rotating system. The nondimensional governing equations are:

$$\begin{gathered}
R(f'f'' - f''') = f^{(4)} + \nu k', \\
R(f'k - fk') = k'' - \nu f', \\
R(gf' - fg') = g'' + \nu h + B, \\
R(gk - fh') = h'' - \nu g, \\
f(0) = -1, \qquad f(1) = -\beta, \qquad f'(0) = f'(1) = 0, \\
k(0) = k(1) = g(0) = g(1) = h(0) = h(1) = 0.
\end{gathered}$$

f, g, h, and k describe the flow, and ν, R, B, β are parameters. There are boundary layers at both 0 and 1 as well as internal boundary layers, which makes this problem extremely difficult. For ν and R small, the homotopy map

$$\rho_a(\lambda, v) = \lambda F(v) + (1 - \lambda)(v - a),$$

where v consists of the appropriate initial conditions and $F(v)$ is defined by shooting, was adequate to solve the problem. Newton and quasi-Newton methods were completely inadequate for this problem. For $\nu, R \geq 30$, $\beta < 0$, $B = 0.5$ shooting becomes impossible because of the sensitivity of the problem, and $F(v)$ defined by a finite difference approximation of the two-point boundary value problem was used in the above homotopy. This approach was quite successful [67], even though the resulting $F(v)$ is a high-dimensional nonlinear function.

4.7. Micropolar flow past porous sheets. Eringen [9], [10] introduced the concept of micropolar fluids to provide a mathematical model for the behavior of fluids which exhibit certain microscopic effects arising from the local structure and micro motions of the fluid elements, such as polymeric fluids, liquid crystals and animal blood. Following Eringen [9], [10], the basic equations of motion and continuity for steady two dimensional flow of micropolar fluids in rectangular Cartesian coordinates xyz with the velocity vector $\vec{U} = [\,u(x,y),\, v(x,y),\, 0\,]$ and the microrotation vector $\vec{\sigma} = [\,0,\, 0,\, \sigma\,(x,y)\,]$ are

$$\rho\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) = -\frac{\partial p}{\partial x} + (\mu + k)\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + k\frac{\partial \sigma}{\partial y} \tag{108}$$

$$\rho\left(u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right) = -\frac{\partial p}{\partial y} + (\mu + k)\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + k\frac{\partial \sigma}{\partial x} \tag{109}$$

$$J\rho\left(u\frac{\partial \sigma}{\partial x} + v\frac{\partial \sigma}{\partial y}\right) = \gamma\left(\frac{\partial^2 \sigma}{\partial x^2} + \frac{\partial^2 \sigma}{\partial y^2}\right) - 2k\sigma + k\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \tag{110}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{111}$$

where ρ, μ, p are the density, viscosity, and pressure, respectively. Further γ is the microrotational coupling coefficient, k is the microrotational diffusivity, and J is the square of a length typical of the microstructure. Eringen [10] showed through thermodynamic arguments that γ, k and μ are all non-negative. Equations (108)–(110) reduce to the Navier-Stokes equations for steady two dimensional flow of incompressible Newtonian fluids when $\gamma = k = J = 0$. The velocity

and microrotation are uncoupled when $k = 0$ and the macroscopic motion is unaffected by the microrotations.

We consider the flow past a wall coinciding with the plane $y = 0$, with the flow confined to $y > 0$, except for injection (or suction) through the wall. Keeping the origin fixed, the wall is stretched by introducing two equal and opposite forces along the x-axis. With the usual boundary layer assumptions Equations (108)–(111) reduce to the following form:

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \nu\frac{\partial^2 u}{\partial y^2} + K\frac{\partial \sigma}{\partial y} \tag{112}$$

$$\tilde{J}\left(u\frac{\partial \sigma}{\partial x} + v\frac{\partial \sigma}{\partial y}\right) = G\frac{\partial^2 \sigma}{\partial y^2} - 2\sigma - \frac{\partial u}{\partial y} \tag{113}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{114}$$

with $\nu = \dfrac{\mu + k}{\rho}$, $K = \dfrac{k}{\rho}$, $G = \dfrac{\gamma}{k}$, and $\tilde{J} = \dfrac{J\rho}{k}$. The boundary conditions are

$$u = Cx, \quad v = -\sqrt{\nu C}\,A, \quad \sigma = 0 \qquad \text{at} \quad y = 0 \tag{115}$$

$$u \longrightarrow 0, \quad \sigma \longrightarrow 0 \qquad \text{as} \quad y \longrightarrow \infty \tag{116}$$

where $C > 0$, $A > 0$ corresponds to suction, and $A < 0$ corresponds to injection.

Using the transformations

$$u = Cx\,f_1'(\eta) \tag{117}$$

$$v = -\sqrt{\nu C}\,f_1(\eta) \tag{118}$$

$$\sigma = \sqrt{C^3/\nu}\,x\,f_2(\eta) \tag{119}$$

$$\eta = \sqrt{C/\nu}\,y \tag{120}$$

Equations (112)–(114) reduce to the following form:

$$f_1''' = -f_1 f_1'' + (f_1')^2 - C_1 f_2' \tag{121}$$

$$f_2'' = C_2^{-1}\left[f_1'' + 2f_2 - C_3\left(f_1 f_2' - f_1' f_2\right)C_1^{-1}\right] \tag{122}$$

where f_1 denotes $f_1(\eta)$, f_2 denotes $f_2(\eta)$, and the non-dimensional parameters

$$C_1 = \frac{K}{\nu}, \quad C_2 = \frac{GC}{\nu}, \quad \text{and} \quad C_3 = \frac{\tilde{J}C}{\nu}. \tag{123}$$

The boundary conditions corresponding to (115) and (116) are

$$f_1 = A, \quad f_1' = 1, \quad f_2 = 0 \qquad \text{for} \quad \eta = 0 \tag{124}$$

$$f_1' \longrightarrow 0, \quad f_2 \longrightarrow 0 \qquad \text{as} \quad \eta \longrightarrow \infty. \tag{125}$$

Define

$$\mathbf{V} = \begin{pmatrix} f_1''(0) \\ f_2'(0) \end{pmatrix} \tag{126}$$

and let $f_1(\eta;\mathbf{V})$, $f_2(\eta;\mathbf{V})$ denote the solution of the initial value problem given by (121) and (122) with the initial conditions (124) and (126). Now note that the original two point boundary value problem (121)–(125) is numerically equivalent to solving the nonlinear system of equations

$$\mathbf{F}(\mathbf{V}) = \begin{pmatrix} f_1'(\tau;\mathbf{V}) \\ f_2(\tau;\mathbf{V}) \end{pmatrix} = 0 \tag{127}$$

where τ is chosen large enough so that $|f_1(\eta) - f_1(\tau)| < \epsilon$ and $|f_2(\eta)| < \epsilon$ for $\tau \leq \eta < \infty$ and a given $\epsilon > 0$.

Algorithms for solving the nonlinear system (127) typically require partial derivatives such as $\frac{\partial f_2}{\partial \mathbf{V}_k}$. These derivatives are calculated by an approach similar to that described in Section 4.4. The nonlinear system of equations (127) can be solved by a globally convergent homotopy method using the homotopy map (2). Such methods are necessary for highly nonlinear problems like (127) since locally convergent methods like Newton's method diverge unless the starting point is very close to the solution and quasi-Newton methods frequently converge to spurious solutions. Considerable computational experience with nonlinear systems of equations arising from fluid mechanics problems indicates that such globally convergent methods are indeed necessary, unless, of course, one is willing to solve a large number of nonlinear systems varying the parameters slowly.

Quasi-Newton methods, such as those implemented in Argonne National Laboratory's MINPACK [21] subroutine package, are robust and usually much more efficient than a globally convergent homotopy method. However, quasi-Newton methods frequently fail by converging to spurious solutions of $\mathbf{F}(\mathbf{V}) = 0$; that is, there are critical points of $\mathbf{F}(\mathbf{V})^t\mathbf{F}(\mathbf{V})$ which fail to satisfy $\mathbf{F}(\mathbf{V}) = 0$. Hence a reasonable overall strategy is to try an inexpensive quasi-Newton algorithm first; and, if that fails, then resort to the expensive but guaranteed homotopy algorithm. To obtain solutions for the fluid flow past a *porous* stretching sheet with suction parameter A, the homotopy method was used to solve for $A = 0$. These solutions produced close initial estimates for $A \neq 0$ as A moved in either direction from 0. Thus, since initial starting values were good, the quasi-Newton algorithm was generally successful for small τ, however, it failed for large (> 10) values of τ (corresponding to $|A| > 2$) and therefore required the solution to a number of nonlinear systems until a sufficiently large τ was obtained.

For large τ (say $\tau > 15$) the quasi-Newton method, when it converged, took less than a minute of CPU time on a VAX 11/780 to solve (127). The homotopy method for a similar problem sometimes took over 10 minutes of CPU time, but never failed to converge [11].

4.8. Magnetohydrodynamic (MHD) flow and heat transfer. This example concerns the magnetohydrodynamic flow and heat transfer about a rotating disk with suction and injection at the disk surface. Some important applications are boundary layer control, cooling of turbine blades, and cooling the skins of high speed aircraft. Another significant application is to model the boundary layer on the face of a crystal grown by the Czochralski method with an axial magnetic field. The goal is to describe the effects of an axial magnetic field and suction (or injection) on the flow and heat transfer about an insulated rotating disk. When the disk is conducting, adding a magnetic field promotes the motion of the fluid, whereas if the disk is insulated, adding a magnetic field decreases the flow velocities.

Let the disk lie in the plane $z = 0$ and the space $z \geq 0$ be occupied by a homogeneous, incompressible, electrically conducting viscous fluid. Here (r,θ,z) are cylindrical coordinates, B_0 is the externally applied magnetic field in the z direction, ω is the angular velocity of the disk, T_w is

the uniform temperature at the disk surface and T_∞ is the ambient fluid temperature. The basic equations for a nonconducting disk are modified to include the Lorenz force $\vec{J} \times \vec{B}$ ($\vec{J}$ being the current density).

In cylindrical coordinates (r,θ,z), assuming angular symmetry, the equations of motion are

$$\rho\left[\left(u\frac{\partial}{\partial r}+w\frac{\partial}{\partial z}\right)u-\frac{v^2}{r}\right]=-\frac{\partial p}{\partial r}+\mu\left(\nabla^2 u-\frac{u}{r^2}\right)-\sigma u B_0^2, \tag{128}$$

$$\rho\left[\left(u\frac{\partial}{\partial r}+w\frac{\partial}{\partial z}\right)v+\frac{uv}{r}\right]=\mu\left(\nabla^2 v-\frac{v}{r^2}\right)-\sigma v B_0^2, \tag{129}$$

$$\rho\left[u\frac{\partial}{\partial r}+w\frac{\partial}{\partial z}\right]w=-\frac{\partial p}{\partial z}+\mu\nabla^2 w, \tag{130}$$

and the continuity equation is

$$\frac{\partial}{\partial r}(ru)+\frac{\partial}{\partial z}(rw)=0, \tag{131}$$

where u, v and w are the velocity components in the r, θ, z directions respectively, ρ is the density of the fluid, μ is the coefficient of viscosity, p is the pressure, σ is the electrical conductivity and

$$\nabla^2=\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial z^2}.$$

Further, equations (128)–(130) assume that the induced electric field is negligible compared with the imposed magnetic field. This assumption is valid for flow at low magnetic Prandtl number. The energy equation is

$$\left(u\frac{\partial}{\partial r}+w\frac{\partial}{\partial z}\right)T=\alpha\nabla^2 T, \tag{132}$$

in which T is the static temperature and α is the thermal diffusivity. The boundary conditions of the problem are

$$\left.\begin{aligned} u&=0\\ v&=r\omega\\ w&=-H_w\\ T&=T_w \end{aligned}\right\}\ \text{at } z=0, \qquad \left.\begin{aligned} u&\to 0\\ v&\to 0\\ T&\to T_\infty \end{aligned}\right\}\ \text{as } z\to\infty, \tag{133}$$

where $H_w > 0$ corresponds to suction and $H_w < 0$ corresponds to injection.

Introduce the following relations

$$\begin{aligned} u&=r\omega F(\eta), \qquad v=r\omega G(\eta), \qquad w=(\omega\nu)^{1/2}H(\eta),\\ P(\eta)&=\frac{p}{\mu\omega}, \qquad \Theta(\eta)=\frac{T-T_\infty}{T_w-T_\infty}, \qquad \eta=z\left(\frac{\omega}{\nu}\right)^{1/2}. \end{aligned} \tag{134}$$

Also, from equation (131) we have

$$F(\eta)=-\frac{H'(\eta)}{2}. \tag{135}$$

Substituting equations (134) and (135) into equations (128)–(132) yields

$$H''' = HH'' - \frac{(H')^2}{2} + mH' + 2G^2, \tag{136a}$$

$$G'' = HG' - H'G + mG, \tag{136b}$$

$$\Theta'' = PrH\Theta', \tag{137}$$

where prime denotes differentiation with respect to η. The Prandtl number Pr and the magnetic parameter m are given by

$$\mathrm{Pr} = \frac{\nu}{\alpha}, \qquad m = \frac{\sigma B_0^2}{\rho\omega}, \tag{138}$$

where $\nu = \mu/\rho$ is the kinematic coefficient of viscosity. In terms of the new variables defined in (134), the new boundary conditions from (133) are

$$\left.\begin{aligned} H' &= 0 \\ H &= -A \\ G &= 1 \\ \Theta &= 1 \end{aligned}\right\} \text{ at } \eta = 0, \qquad \left.\begin{aligned} H' &\to 0 \\ G &\to 0 \\ \Theta &\to 0 \end{aligned}\right\} \text{ as } \eta \to \infty, \tag{139}$$

where A is the nondimensional velocity normal to the disk surface. $A > 0$ represents suction while $A < 0$ represents injection. Observe that (137) decouples from (136a) and (136b), and that once $H(\eta)$ has been determined, $\Theta(\eta)$ can be computed by solving a relatively easy one-dimensional two-point boundary value problem.

In practice the infinite boundary conditions in (139) are replaced by

$$H'(\tau) = G(\tau) = \theta(\tau) = 0 \tag{139a}$$

for some τ sufficiently large such that

$$|H(\eta) - H(\tau)| + |G(\eta)| + |\theta(\eta)| \approx 0 \quad \text{for } \tau \le \eta < \infty.$$

Let S_n be the finite dimensional vector space with basis $\{B_{j,k,\mathbf{t}}(x)\}_{j=1}^{n}$, where $B_{j,k,\mathbf{t}}(x)$ is the jth B-spline of order k (degree $\le k-1$) defined on the knot sequence $\mathbf{t} = (t_1, t_2, \ldots, t_{n+k})$. When there is no ambiguity $B_{j,k,\mathbf{t}}(x)$ is simply written as $B_j(x)$. For this problem the knot sequence $\mathbf{t}$ is based on the breakpoint sequence

$$\begin{aligned} \Xi = (&0, .25, .50, .75, 1.0, 1.25, 1.50, 1.75, 2.0, 2.25, 2.50, 2.75, \\ &3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0, 9.0, 11.0, 13.0, 15.0, \\ &18.0, 21.0, 24.0, 28.0, 32.0, 36.0, 41.0, 46.0, 51.0, 60.0, \\ &70.0, 80.0, 90.0, 100.0), \end{aligned}$$

following the convention

$$t_1 = t_2 = \cdots = t_k \qquad \text{and} \qquad t_{n+1} = t_{n+2} = \cdots = t_{n+k}$$

used by deBoor [6]. These repeated knots essentially mean that the spline is free at the endpoints of the approximation interval $[t_k, t_{n+1}]$. Note the distinction between knots and breakpoints. The rest of the knots are simple, i.e., $t_i < t_{i+1}$ for $i = k, \ldots, n$. The functions $H(\eta)$ and $G(\eta)$ have boundary layers (large derivatives) near $\eta = 0$, and then asymptotically approach a constant as $\eta \to \infty$. The appropriate value for τ (unknown beforehand) is where the solutions level off (within some error tolerance) at their asymptotic value. Short of dynamically adapting the knot sequence (which, as deBoor [6] points out, is rarely cost effective), a reasonable strategy is to space the knots farther apart as η increases. Depending on the values of n and k, only an initial subsequence of the breakpoint sequence Ξ above is used.

The approximations are

$$H(\eta) = \sum_{j=1}^{N+2} \alpha_j B_j(\eta), \tag{140}$$

$$\alpha_1 = -A, \quad \alpha_2 = \frac{A\, B_1'(0)}{B_2'(0)}, \quad \alpha_{N+2} = \frac{-\alpha_{N+1} B_{N+1}'(\tau-)}{B_{N+2}'(\tau-)}, \tag{141}$$

$$G(\eta) = \sum_{j=1}^{N+2} \beta_j B_j(\eta), \tag{142}$$

$$\beta_1 = 1, \quad \beta_{N+2} = 0. \tag{143}$$

The boundary conditions (139) force the equations (141) and (143). The Galerkin approximation is the nonlinear system of equations

$$\begin{aligned} \langle -H''' + H\, H'' - (H')^2/2 + m\, H' + 2\, G^2,\ B_i \rangle &= 0, \qquad i = 3, \ldots, N+1, \\ \langle -G'' + H\, G' - H'\, G + m\, G,\ B_i \rangle &= 0, \qquad i = 2, \ldots, N+1, \end{aligned} \tag{144}$$

where

$$\langle u, v \rangle = \int_0^\tau u(\eta) v(\eta)\, d\eta.$$

Let $Y = (\alpha_3, \alpha_4, \ldots, \alpha_{N+1}, \beta_2, \beta_3, \ldots, \beta_{N+1})^t$ and $F(Y) = 0$ be given by the $p = 2N - 1 = 2n - 5$ equations (144).

This MHD problem is fairly difficult, and both homotopy and standard quasi-Newton methods based on simple shooting and multiple shooting are known to fail for certain ranges of the parameters A and m. Finite difference, collocation, or Galerkin formulations handle the fluid dynamics boundary layers better, but the failure of quasi-Newton algorithms (started distant from the solution) on such large dimensional approximations to fluid dynamics problems is well documented [73]. Thus the alternatives are to combine a homotopy algorithm with a finite difference approximation, a collocation approximation, or a Galerkin approximation (the present approach).

Table 2 shows some numerical results obtained by applying subroutine FIXPNF (normal flow algorithm) of HOMPACK to (144). The normal flow algorithm was the most efficient of the three algorithms in HOMPACK for this problem. While predicting the best algorithm for a given problem is risky, some general rules of thumb are: the ODE-based algorithm is the most robust but the most expensive; if the zero curve γ has very sharp turns (it doesn't for this problem) then the ODE-based algorithm is the best; if the Jacobian matrices are very expensive to evaluate (they aren't for this problem) then the quasi-Newton augmented Jacobian scheme is best; otherwise the

TABLE 2

A	m	k	$N+2$	$-H(\tau)$	$-H(\infty)$	NFE	CPU time	arc length
−1.0	1.0	6	12	−.80700	−.43166	29	13:28	1.309
−1.0	2.0	6	12	−.88173	−.78156	29	13:28	1.387
−1.0	4.0	6	12	−.94477	−.93015	30	14:37	1.518
0.0	1.0	6	12	.11991	.25331	25	11:13	1.067
0.0	2.0	6	12	.07372	.10858	25	11:24	1.103
0.0	4.0	6	12	.03558	.04078	20	9:07	1.198
1.0	1.0	6	12	1.06079	1.0898	21	9:38	1.141
1.0	2.0	6	12	1.03959	1.0481	20	8:57	1.187
1.0	4.0	6	12	1.02103	1.0225	15	6:42	1.272
2.0	1.0	6	12	2.02744	2.0318	18	8:00	1.257
2.0	2.0	6	12	2.01968	2.0213	24	10:38	1.296
2.0	4.0	6	12	2.01193	2.0123	15	6:37	1.365
4.0	1.0	6	12	4.00625	4.0064	18	7:58	1.471
4.0	2.0	6	12	4.00530	4.0054	18	7:58	1.491
4.0	4.0	6	12	4.00402	4.0041	15	6:38	1.530
−1.0	1.0	6	24	−.43877	−.43166	32	1:09:37	1.916
−1.0	2.0	6	24	−.78196	−.78156	27	58:20	1.604
−1.0	4.0	6	24	−.93019	−.93015	24	51:32	1.953
0.0	1.0	6	24	.25286	.25331	26	55:45	1.986
0.0	2.0	6	24	.10852	.10858	21	44:53	1.804
0.0	4.0	6	24	.04073	.04078	25	53:27	1.903
−1.0	1.0	6	32	−.43165	−.43166	33	2:15:48	2.765
−1.0	2.0	6	32	−.78158	−.78156	34	2:20:28	2.334
−1.0	4.0	6	32	−.93018	−.93015	27	1:52:50	2.752
−1.0	1.0	4	24	−.42854	−.43166	41	55:37	2.196
−1.0	2.0	4	24	−.78043	−.78156	38	51:12	1.810
−1.0	4.0	4	24	−.93062	−.93015	29	39:06	2.128

normal flow algorithm is best. The values $H(\infty)$ are from [16], and τ can be inferred from n, k, and the breakpoint sequence Ξ listed above. The integrals in (144) were computed by 10-point Gaussian quadrature over each subinterval, and are thus essentially exact. NFE is the number of Jacobian matrix evaluations, and the format of the CPU time (on a VAX 11/780) is hh:mm:ss. The local curve tracking tolerance was 10^{-4} and the final accuracy (the end game tolerance) was 10^{-8}.

The problem gets easier (weaker boundary layers, asymptotic value reached sooner) as either A or m increases. For example, with $n = N+2 = 12$ and spline order $k = 6$, $\tau = t_{13} = t_{18} = 1.75$, and as the plots in [16] show, this τ is clearly not large enough for $A = -1$, $m = 1.0$, but it is about right for $A = m = 4$. Taking $n = 24$, $k = 6$ gives $\tau = 7.0$, which produces better results for the case $A = -1$, $m = 1.0$, and finally $n = 32$, $k = 6$ gives $\tau = 24.0$ and the theoretically best possible results ($\mathcal{O}(h^6) = \mathcal{O}\big((1/4)^6\big) = \mathcal{O}(10^{-4})$) for $A = -1$, $m = 1.0$. The lower order spline ($k = 4$) produces the expected $\mathcal{O}(h^4)$ accuracy at the right τ for $A = -1$, $m = 4.0$, but not for $A = -1$, $m = 1.0$ because τ is too small for this case.

There are three concluding observations. First, although the theory in §2 is not directly applicable to this MHD problem, these numerical results suggest that the homotopy algorithm is more widely applicable than the theory indicates. Second, the accuracy is exactly what would

be expected of a Galerkin approximation given the spline order and knot spacing. Third, all the solutions for Table 2 were obtained on the first computer run by HOMPACK with no tweaking of initial points or error tolerances whatsoever.

5. References.

[1] E. Allgower and K. Georg, *Simplicial and continuation methods for approximating fixed points*, SIAM Rev., 22 (1980), pp. 28–85.

[2] J.-F. M. Barthelemy and M. F. Riley, *Improved multilevel optimization approach for the design of complex engineering systems*, AIAA J., 26 (1988), pp. 353–360.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[4] S. C. Billups, *An augmented Jacobian matrix algorithm for tracking homotopy zero curves*, M.S. Thesis, Dept. of Computer Sci., VPI & SU, Blacksburg, VA, Sept., 1985.

[5] S. N. Chow, J. Mallet-Paret, and J. A. Yorke, *Finding zeros of maps: Homotopy methods that are constructive with probability one*, Math. Comput., 32 (1978), pp. 887–899.

[6] C. deBoor, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.

[7] J. E. Dennis and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[8] G. H. Ellis and L. T. Watson, *A parallel algorithm for simple roots of polynomials*, Comput. Math. Appl., 10 (1984), pp. 107–121.

[9] A. C. Eringen, *Simple microfluids*, Int. J. Engng. Sci., 2 (1964), pp. 205–217.

[10] A. C. Eringen, *Theory of micropolar fluids*, J. Math. Mech., 16 (1966), pp. 1–18.

[11] M. W. Heruska, L. T. Watson, and K. K. Sankara, *Micropolar flow past a porous stretching sheet*, Comput. & Fluids, 14 (1986), pp. 117–129.

[12] M. P. Kamat, L. T. Watson, and V. B. Venkayya, *A quasi-Newton versus a homotopy method for nonlinear structural analysis*, Comput. & Structures, 17 (1983), pp. 579–585.

[13] H. B. Keller, *Numerical Solution of Two-point Boundary Value Problems*, Society for Industrial and Applied Mathematics, Philadelphia, 1976.

[14] G. Kreisselmeier and R. Steinhauser, *Systematic control design by optimizing a vector performance index*, Proc. IFAC Symp. on Computer Aided Design of Control Systems, Zurich, Switzerland, (1979), pp. 113–117.

[15] M. Kubicek, *Dependence of solutions of nonlinear systems on a parameter*, ACM Trans. Math. Software, 2 (1976), pp. 98–107.

[16] S. K. Kumar, W. I. Thacker, and L. T. Watson, *Magnetohydrodynamic flow and heat transfer about a rotating disk with suction and injection at the disk surface*, Comput. & Fluids, 16 (1988), pp. 183–193.

[17] S. K. Kumar, W. I. Thacker, and L. T. Watson, *Magnetohydrodynamic low past a porous rotating disk in a circular magnetic field*, Internat. J. Numer. Methods Fluids, 8 (1988), pp. 659–669.

[18] S. K. Kumar, W. I. Thacker, and L. T. Watson, *Magnetohydrodynamic flow between a solid rotating disk and a porous stationary disk*, Appl. Math. Modelling, to appear.

[19] S. K. Kumar, W. I. Thacker, and L. T. Watson, *Rotating magnetohydrodynamic flow about a stationary disk in a circular magnetic field*, Comput. Math. Appl., to appear.

[20] H. H. KWOK, M. P. KAMAT, AND L. T. WATSON, *Location of stable and unstable equilibrium configurations using a model trust region quasi-Newton method and tunnelling*, Comput. & Structures, 21 (1985), pp. 909–916.

[21] O.L. MANGASARIAN, *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Appl. Math., 31 (1976), pp. 89–92.

[22] R. MEJIA, *CONKUB: A conversational path-follower for systems of nonlinear equations*, J. Comput. Phys., 63 (1986), pp. 67–84.

[23] J. J. MORÉ, B. S. GARBOW, AND K. E. HILLSTROM, *User Guide for MINPACK-1*, ANL-80-74, Argonne National Laboratory, (1980).

[24] A. P. MORGAN, *A transformation to avoid solutions at infinity for polynomial systems*, Appl. Math. Comput., 18 (1986), pp. 77–86.

[25] ——, *A homotopy for solving polynomial systems*, Appl. Math. Comput., 18 (1986), pp. 87–92.

[26] ——, *Solving polynomial systems using continuation for engineering and scientific problems*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[27] A. P. MORGAN AND L. T. WATSON, *A globally convergent parallel algorithm for zeros of polynomial systems*, Nonlinear Anal., to appear.

[28] ——, *Solving nonlinear equations on a hypercube*, in Super and Parallel Computers and Their Impact on Civil Engineering, M. P. Kamat (ed.), ASCE Structures Congress '86, New Orleans, LA, 1986, pp. 1–15.

[29] W. PELZ AND L. T. WATSON, *Message length effects for solving polynomial systems on a hypercube*, Parallel Comput., 10 (1989), pp. 161–176.

[30] A. B. POORE AND D. SORIA, *Continuation algorithms for linear programming*, in preparation.

[31] A. B. POORE AND Q. AL-HASSAN, *The expanded Lagrangian system for constrained optimization problems*, SIAM J. Control Optim., 26 (1988), pp. 417–427.

[32] W. C. RHEINBOLDT AND J. V. BURKARDT, *Algorithm 596: A program for a locally parameterized continuation process*, ACM Trans. Math. Software, 9 (1983), pp. 236-241.

[33] H. SCARF, *The Computation of Economic Equilibria*, Yale University Press, (1973).

[34] L. F. SHAMPINE AND M. K. GORDON, *Computer Solution of Ordinary Differential Equations: The Initial Value Problem*, W. H. Freeman, San Francisco, 1975.

[35] Y. S. SHIN, R. T. HAFTKA, L. T. WATSON, AND R. H. PLAUT, *Tracing structural optima as a function of available resources by a homotopy method*, Comput. Methods Appl. Mech. Engrg., 70 (1988), pp. 151–164.

[36] ——, *Design of laminated plates for maximum buckling load*, J. Composite Materials, 23 (1989), pp. 348–369.

[37] G. VASUDEVAN, L. T. WATSON, AND F. H. LUTZE, *A homotopy approach for solving constrained optimization problems*, Tech. Rep. 88–50, Dept. of Computer Sci., VPI&SU, Blacksburg, VA, 1988.

[38] C. Y. WANG AND L. T. WATSON, *Squeezing of a viscous fluid between elliptic plates*, Appl. Sci. Res., 35 (1979), pp. 195–207.

[39] ——, *Viscous flow between rotating discs with injection on the porous disc*, Z. Angew. Math. Phys., 30 (1979), pp. 773–787.

[40] ——, *On the large deformations of C-shaped springs*, Internat. J. Mech. Sci., 22 (1980), pp. 395–400.

[41] ——, *Theory of the constant force spring*, J. Appl. Mech., 47 (1980), pp. 956–958.
[42] ——, *The fluid-filled cylindrical membrane container*, J. Engrg. Math., 15 (1981), pp. 81–88.
[43] ——, *Equilibrium of heavy elastic cylindrical shells*, J. Appl. Mech., 48 (1981), pp. 582–586.
[44] ——, *The elastic catenary*, Internat. J. Mech. Sci., 24 (1982), pp. 349–357.
[45] ——, *Free rotation of a circular ring about a diameter*, Z. Angew. Math. Phys., 34 (1983), pp. 13–24.
[46] C. Y. Wang, L. T. Watson, and M. P. Kamat, *Buckling, postbuckling and the flow through a tethered elastic cylinder under external pressure*, J. Appl. Mech., 50 (1983), pp. 13–18.
[47] L. T. Watson, S. C. Billups, and A. P. Morgan, *HOMPACK: A suite of codes for globally convergent homotopy algorithms*, Tech. Rep. 85–34, Dept. of Industrial and Operations Eng., Univ. of Michigan, Ann Arbor, MI, 1985, and ACM Trans. Math. Software, 13 (1987), pp. 281–310.
[48] L. T. Watson, J. P. Bixler, and A. B. Poore, *Continuous homotopies for the linear complementarity problem*, Tech. Report TR–87–38, Dept. of Computer Sci., VPI&SU, Blacksburg, VA, 1987 and SIAM J. Matrix Anal. Appl., 10 (1989), pp. 259–277.
[49] L. T. Watson and D. Fenner, *Chow-Yorke algorithm for fixed points or zeros of C^2 maps*, ACM TOMS, 6 (1980), pp. 252–260.
[50] L. T. Watson and R. T. Haftka, *Modern homotopy methods in optimization*, Comput. Methods Appl. Mech. Engrg., 74 (1989), pp. 289–305.
[51] L. T. Watson, S. M. Holzer, and M. C. Hansen, *Tracking nonlinear equilibrium paths by a homotopy method*, Nonlinear Anal., 7 (1983), pp. 1271–1282.
[52] L. T. Watson, M. P. Kamat, and M. H. Reaser, *A robust hybrid algorithm for computing multiple equilibrium solutions*, Engrg. Comput., 2 (1985), pp. 30–34.
[53] L. T. Watson, T. Y. Li and C. Y. Wang, *Fluid dynamics of the elliptic porous slider*, J. Appl. Mech., 45 (1978), pp. 435–436.
[54] L. T. Watson, K. K. Sankara, and L. C. Mounfield, *Deceleration of a porous rotating disk in a viscous fluid*, Internat. J. Engrg. Sci., 23 (1985), pp. 131–137.
[55] L. T. Watson and L. R. Scott, *Solving Galerkin approximations to nonlinear two-point boundary value problems by a globally convergent homotopy method*, SIAM J. Sci. Stat. Comput., 8 (1987) 768–789.
[56] L. T. Watson and M. R. Scott, *Solving spline-collocation approximations to nonlinear two-point boundary-value problems by a homotopy method*, Appl. Math. Comput., 24 (1987) 333–357.
[57] L. T. Watson and C. Y. Wang, *A homotopy method applied to elastica problems*, Internat. J. Solids Structures, 17 (1981), pp. 29–37.
[58] ——, *Deceleration of a rotating disc in a viscous fluid*, Phys. Fluids, 22 (1979), pp. 2267–2269.
[59] ——, *The circular leaf spring*, Acta Mech., 40 (1981), pp. 25–32.
[60] ——, *Hanging an elastic ring*, Internat. J. Mech. Sci., 23 (1981), pp. 161–168.
[61] ——, *Overhang of a heavy elastic sheet*, Z. Angew. Math. Phys., 33 (1982), pp. 17–23.
[62] ——, *Periodically supported heavy elastic sheet*, J. Engrg. Mech., 109 (1983), pp. 811–820.
[63] ——, *The equilibrium states of a heavy rotating column*, Internat. J. Solids Structures, 19 (1983), pp. 653–658.

[64] L. T. WATSON AND W. H. YANG, *Optimal design by a homotopy method*, Applicable Anal., 10 (1980), pp. 275–284.

[65] ——, *Methods for optimal engineering design problems based on globally convergent methods*, Comput. & Structures, 13 (1981), pp. 115–119.

[66] L. T. WATSON, *A globally convergent algorithm for computing fixed points of C^2 maps*, Appl. Math. Comput., 5 (1979), pp. 297–311.

[67] ——, *Numerical study of porous channel flow in a rotating system by a homotopy method*, J. Comput. Appl. Math., 7 (1981), pp. 21–26.

[68] ——, *Computational experience with the Chow-Yorke algorithm*, Math. Programming, 19 (1980), pp. 92–101.

[69] ——, *Fixed points of C^2 maps*, J. Comput. Appl. Math., 5 (1979), pp. 131–140.

[70] ——, *Solving the nonlinear complementarity problem by a homotopy method*, SIAM J. Control Optim., 17 (1979), pp. 36–46.

[71] ——, *An algorithm that is globally convergent with probability one for a class of nonlinear two-point boundary value problems*, SIAM J. Numer. Anal., 16 (1979), pp. 394–401.

[72] ——, *Solving finite difference approximations to nonlinear two-point boundary value problems by a homotopy method*, SIAM J. Sci. Stat. Comput., 1 (1980), pp. 467–480.

[73] ——, *Engineering applications of the Chow-Yorke algorithm*, Appl. Math. Comput., 9 (1981), pp. 111–133.

[74] ——, *Numerical linear algebra aspects of globally convergent homotopy methods*, Tech. Report TR–85–14, Dept. of Computer Sci., VPI&SU, Blacksburg, VA, 1985, and SIAM Rev., 28 (1986), pp. 529–545.

[75] ——, *Globally convergent homotopy methods: a tutorial*, Appl. Math. Comput., 31BK (1989), pp. 369–396.